IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

WANG

Application No.: 09/870,353

Filed: May 30, 2001

For: IMPROVED NUCLEIC ACID MODIFYING ENZYMES

Examiner: Richard Hutson

Technology Center/Art Unit: 1652

RULE 132 DECLARATION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

I, Dr. Peter Vander Horn, being duly warned that willful false statements and the like are punishable by fine or imprisonment or both, under 18 U.S.C. § 1001, and may jeopardize the validity of the patent application or any patent issuing thereon, state and declare as follows:

1. All statements herein made of my own knowledge are true and statements made on information or belief are believed to be true. The Exhibits (1 - 10) attached hereto are incorporated herein by reference.

2. I received a Ph.D. in microbiology from Cornell University in 1991. A copy of my curriculum vitae is attached as **Exhibit 1**.

3. I am presently employed by MJ Bioworks, Inc. as Vice President of Research, Development, and Engineering. I am primarily responsible for supervising research teams working to improve our scientific instrumentation products. MJ Bioworks is the assignee of the subject patent application.

4. I have read and am familiar with the contents of the application. As I understand the bases for the outstanding rejections, the Examiner believes that the pending claims are overly broad and that it would take undue experimentation to identify members of the genus of non-specific double-stranded nucleic acid binding domains that are either recognized by polyclonal antibodies generated against Sso7d or have at least 50% identity to a 50 amino acid subsequence of Seq. ID No: 2 or a 75% identity to Sac7d.

5. The criteria set forth in the claims was intended to provide us with claim scope that embraced both naturally occurring proteins in the family of non-specific DNA binding Archaeal proteins as well as "Archaeal 7 kDa muteins". By Archaeal 7 kDa muteins, I am referring to man-made recombinantly produced proteins that are derived from naturally occurring proteins. In this context, muteins differ from their parent proteins by the introduction of amino acid changes where those changes do not markedly alter its **DNA binding** properties compared to the parent protein.

6. It is the intent of this declaration to explain in objective scientific reasons, why one of skill can identify working embodiments that fall within the scope of these claims with routine experimentation. In summary, there are three objective reasons and one subjective reason. The three objective reasons are: (i) that genetic variation or drift within the naturally occurring species of Archaeal 7 kDa proteins provides an initial road map for point mutations; (ii) that conventional knowledge of protein chemistry allows for us to predict that biological properties can be preserved so long as amino acid substitutions are conservative in their nature; and (iii) that knowledge of the three dimensional structure of these proteins when bound to DNA permits us to predict areas of non-criticality where substitutions may be freely introduced beyond mere conservative substitutions. As a subjective rationale, we must consider that the family of Archaeal 7 kDa proteins come from extremophilic bacteria that live in acidic environments above the melting temperature of DNA. This group of extremophiles includes many unexplored species that by virtue of their habitats are expected to have Archaeal 7 kDa-like DNA binding proteins. With so many species to be studied and so few cultured it is highly probable that additional members of the family will be discovered with even greater variation than those that are presently known and sequenced.

7. NATURAL VARIATION.

With regard to naturally occurring 7 kDa proteins in the family of Archaeal DNA-binding proteins, there are many family members reported in the literature. It is an accepted convention that proteins with E scores below 0.01 are unlikely to occur by chance and are therefore statistically related. Using Sso7d as a prototype, we studied the family of Archaeal DNA binding proteins reported in GenBank. We noted that there are at least 17 related members of the 7 kDa class of Archaeal proteins. The least related of which has an E value of $9x10^{-6}$.

The evolutionary relationship between the members of this family is made quite clear when you conduct a BlastP search comparing Sso7d to its family members. Using the default parameters provided by the specification on page 16, lines 7-11 with the "Low Complexity" filter set to off to permit us to align the entire 63 amino acids, we get the following results:

		Identity	Similarity
SEQ ID:2.	ATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDAPKELLQMLEKQKK		
1) RNaseP3 of S:	atvkfkykgeekqvdiskikkvwrvgkmisftydegggktgrgavsekdapkellqmmpetgkyfrhklpddypi	90%	95%
2) Sso7d	meismatvkfkykgeekevdiskikkvwrvgkmisftydegggktgrgavsekdapkellqmlekqkk	100%	100%
3) Sso7d	matvkfkykgeekqvdiskikkvwrvgkmisftydegggktgrgavsekdapkellqmlakqkk	98%	100%
4) Sso7d	matvkfkykgeekevdiskikkvwrvgkmisftydegggktgrgavsekdapkellqmlekqkk	100%	100%
5) Sso7d	atvkfkykgeekqvdiskikkvwrvgkmisftydegggktgrgavsekdapkellqmlekqk	98%	100%
6) Sso7d	matvkfkykgeekqvdiskikkvwrvgkmisftydegggktgrgavsekdapkellqmlekqkk	98%	100%
7) Sso7d	atvkfkykgeekevdiskikkvwrvgkmisftydegggktgrgavsekdapkellqmlekqkk	100%	100%
8) Sso7d	atvkfkykgeekqvdiskikkvwrvgkmisftydegggktgrgavsekdapkellqmlekqkk	98%	100%
9) Ssh7B	mvtvkfkykgeekevdtskikkvwrvgkmisftydegggktgrgavsekdapkellqmlekqkk	98%	98%
10) Sso7d mutant	atvkfkykgeekqvdiskikkvwrvgkmisatydegggktgrgavsekdapkellqmlekqk	96%	98%
11) Sso7e/Sto7e	mvtvkfkykgeekevdiskikkvwrvgkmisftydd-ngktgrgavsekdapkellqmleksgkk	91%	93%
12) Sac7a	vkvkfkykgeekevdtskikkvwrvgkmvsftydd-ngktgrgavsekdapkelldmlarae	86%	91%
13) Sac7a/b/d	mvkvkfkykgeekevdtskikkvwrvgkmvsftydd-ngktgrgavsekdapkelldmlaraerekk	81%	90%
14) Sac7e	makvrfkykgeekevdtskikkvwrvgkmvsftydd-ngktgrgavsekdapkelmdmlaraekkk	79%	88%
15) 1SAP/Sac7	kvkfkykgeekevdtskikkvwrvgkmvsftydd-ngktgrgavsekdapkelldmlaraerekk	86%	91%
16) Sac7e	akvrfkykgeekevdtskikkvwrvgkmvsftydd-ngktgrgavsekdapkelmdmlaraekkk	79%	88%
17) Sso Dna binding protein	tvkfkykgeekqvdiskikkvxrvgkmisftydegxgk	92%	94%

From the above BLASTP data, we can see that the natural variation within the family extends to below 80% identity. At a minimum, it was the applicants' intent to encompass in a single claim all naturally occurring known variants of the DNA binding Archaeal protein family. But our knowledge of variants can be extended to include muteins by applying our knowledge of protein chemistry - knowledge that is both routine and predictable in its application.

8. MUTEINS CREATED BY COMBINING NATURALLY OCCURRING VARIATION.

Muteins of Archaeal 7 kDa proteins can be readily created by those of skill exploiting variation within the natural members of the family to create novel combinations of variations. In essence, the naturally occurring members are a road map to defining the critical amino acids from the non-critical amino acids.

A cursory review of the family reveals that the amino and carboxyl termini are not critical to the functionality of these proteins. The amino and carboxyl ends are very tolerant of substitutions and additions. They are sites of divergence between the homologues and the invention. As evidence of the robust nature of these proteins, we placed entire polymerase domains on both the carboxyl and the amino ends without interfering with binding. This was Dr. Wang's rationale for claiming sequence similarity to a 50-amino acid subsequence, rather than to the entire protein. Biological functionality appears to be determined by the conserved amino acids that form the internal core of these proteins (see Choli et al. (1988) *Biochimica et Biophysica Acta*, 950:193-203 at 202) **(Exhibit 2)**. But even there the identity is not 100%.

9. MUTEINS CREATED BY INTRODUCTION OF CONSERVED SUBSTITUTIONS.

In addition to the introducing combinations of naturally occurring variations into a prototype 7 kDa binding protein, those of skill can also substitute *conserved* amino acids for naturally occurring ones that have not been found to vary in nature. Classic examples of such pairings are lysine and arginine, alanine and glycine, glutamine and asparagine, and aspartic acid and glutamic acid. All of which appear in this family of proteins. For example, there are 12 residues of Sso7d 63 residues in which natural variations are known. By substituting conserved amino acids for another 20 residues, we can easily produce a non-specific 7 kDa Archaeal mutein that would almost certainly work to improve processivity of a polymerase.

10. MUTEINS DERIVED FROM STUDIES OF THREE DIMENSIONAL ANALYSES.

We need not limit our muteins to combinations of naturally occurring amino acid variations nor to those that are unnatural but between amino acids of similar chemical properties. This is because the three dimensional structure of these proteins when interacting with DNA is known. See **Exhibit 3** Gao *et al.*

Knowledge of three dimensional features provides yet another strategy permitting protein chemists to engineer away from the native sequences because it provides structural activity relationships between the protein domains and DNA. Knowing which domains play a role in DNA binding and which are non-critical for binding permits us to think beyond mere conservative amino acid substitution and to allow for Archaeal 7 kDa muteins with lower percent identities than if we confined our mutein development strategy to the first two objective approaches.

Attached to this declaration as **Exhibits 4-8** are enlargements of figures derived from the data of Gao, et al. with an accession number of 1BNZ.[1] **Exhibit 4** is a ribbon diagram of the

[1] These figures are derived from the protein crystal coordinates that Gao submitted to the protein structure database. Submission is a requirement

crystal structure of Sso7d bound to DNA. The beta sheets of the protein are in yellow, the alpha helix is in green. Unstructured regions are in blue.

As predicted, the unstructured regions are sites where divergences from Sso7d among the group of related proteins cluster. One skilled in the art could place additional insertions into these sites that will decrease sequence identity in blast analyses. For example, a thermostable loop can be placed in the G37, G38, G39 turn.

In addition, the entire alpha helix (green) is highly mutable. This is evidenced by the fact that a great deal of natural variation of the homologs is observed in this domain. It should be noted that the naturally occurring mutations in this domain do appear to preserve the presence of an alpha helix and this region does not interact with the DNA substrate. Therefore, additional mutations could be introduced into the alpha helix (as long as they preserve the secondary structure) and serve to further lower the amino sequence identity compared to SEQ ID 2.

Using the three dimensional figures, those of skill could also take note that the differences in composition and length between Sso7 and Sac7 proteins cluster in the turns between beta sheets and in amino acids facing away from the DNA binding domain in the crystal structure. So these domains are also areas of plasticity.

The papers cited in the patent application describe several exposed lysine residues that are methylated *in vivo*. These sites are not involved in DNA binding but appear to be regulatory. As our work is independent of bacterial gene regulation, these lysines could be mutated so long as they do not interact with the DNA substrate. As can be seen in **Exhibits 5** though **8**, many of these lysine residues project away from the domain and do not interact with DNA. These residues are excellent candidates for mutagenesis. One skilled in the art would recognize that these could be changed to arginine residues without affecting DNA binding.

I was able to find 10 such sites by examining the crystal structure. **Exhibit 5** shows lysines 19, 40, 49, and 53 projecting away from the DNA binding surface of the protein. **Exhibit 6** also shows lysines 49, 61, and 64. Exhibit 7 shows lysine 63 and Exhibit 8 shows lysines 5 and 13. K to R derivatives already exist for positions 5 and 61, validating this approach. No divergence from the Sso7d sequence has been observed for the remaining 8 lysines, probably because of the regulatory role alluded to earlier. Mutating these lysines can yield an additional 8 differences from SEQ ID No. 2, or 13%.

---

similar to the requirement that sequences be deposited into Genbank with an accession number. The accession code for Sso7d protein bound to DNA is 1BNZ. The coordinates are viewed and turned into these figures using the program Cn3d, which is freely available at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Structure.

For these varied but objective reasons, one skilled in the art could with a combination of conserved substitutions, insertions, deletions, and exchanges of mutable sites construct DNA binding proteins that are very divergent from SEQ ID: 2 and Sac7d. I will discuss specific percentages later in this Declaration.

11. OTHER EXTREMOPHILES WILL HAVE ARCHAEAL 7 kDa LIKE PROTEINS.

Beyond the objective reasons presented above, there is a subjective reason why a percentage below 90% is needed to avoid routine engineering around the presently issued claims. As of today there have been many Archaeal 7 kDa proteins that have already been reported, it should be noted that these proteins are very abundant in *Sulfolobus* species. In fact, they are probably abundant in any organism that has to live in acid at >70°C chemolithotrophically. Here are *S. Solfataricus's* relatives many of which are expected to contain Sso7d-related proteins.

```
Archaea; Crenarchaeota; Thermoprotei; Sulfolobales
Sulfolobaceae
     Acidianus
          Acidianus ambivalens
          Acidianus brierleyi
          Acidianus infernus
          Acidianus tengchongenses
          Metallosphaera
          Metallosphaera prunae
          Metallosphaera sedula
          Metallosphaera sp. GIB11/00
          Metallosphaera sp. J1
          Metallosphaera sp. TA-2
          environmental samples
               uncultured Metallosphaera sp.

     Stygiolobus
          Stygiolobus azoricus
          environmental samples
               uncultured Stygiolobus sp.

     Sulfolobus
          Sulfolobus acidocaldarius
          Sulfolobus islandicus
          Sulfolobus metallicus
          Sulfolobus shibatae
          Sulfolobus solfataricus
          Sulfolobus thuringiensis
          Sulfolobus tokodaii
          Sulfolobus yangmingensis
          Sulfolobus sp.
          Sulfolobus sp. AMP12/99
          Sulfolobus sp. CH7/99
          Sulfolobus sp. FF5/00
          Sulfolobus sp. MV2/99
          Sulfolobus sp. MVSoil3/SC2
```

```
        Sulfolobus sp. MVSoil6/SC1
        Sulfolobus sp. NGB23/00
        Sulfolobus sp. NGB6/00
        Sulfolobus sp. NL8/00
        Sulfolobus sp. NOB8H2
        Sulfolobus sp. RC3
        Sulfolobus sp. RC6/00
        Sulfolobus sp. RCSC1/01
        Sulfolobus sp. RT8-4
        environmental samples
             uncultured Sulfolobus sp.

    Sulfurisphaera
        Sulfurisphaera ohwakuensis
```

So far only *Sulfolobus solfataricus* and *Sulfolobus tokodaii* genomes have been sequenced.

Given the range of divergence in Archaeal 7 kDa DNA binding proteins set forth above from a tiny portion of species sequenced, it will be trivial to find additional species of these DNA binding proteins that will have 70% or less homology to the presently known prototypes.

12. THE 90% LIMITATION OF THE '424 PATENT INVITES THOSE OF SKILL TO ENGINEER AROUND THE CLAIMS WITH EASE.

Let's look more specifically at the information that was available prior to filing the subject application. Dr. Wang's earlier patent US Pat. No. 6,627,424 ['424] issued with claims covering 90% identity to Sso7d and identity to Sac7d. Below I have created a paired table comparing the relative homology between Sso7d and Sac7d and Sac7d and Sac7e.

As you can see, close relatives of Sso7d, (i.e., Sac7a,b,d and e) are not covered by the recited percentage in our '424 patent claims. But a pair-wise alignment of these sequences to the two specific examples gives one a clear road map to implementing the invention with any of the naturally occurring homologues.

```
Sso7d alignment to Sac7d.
Sso7d: 1  MATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDAPKELLQML---EKQKK 64  Identity Similarity
          M--VKFKYKGEEKEVD-SKIKKVWRVGKM+SFTYD+--GKTGRGAVSEKDAPKELL-ML---E++KK      80%      85%
Sac7d: 1  MVKVKFKYKGEEKEVDTSKIKKVWRVGKMVSFTYDD-NGKTGRGAVSEKDAPKELLDMLARAEREKK 66

Note: the percent identity changes to 82% and the similarity changes to 88% if Seq ID 2 is used. This is because Seq
ID 2 is Sso7d without the MET. One skilled in the art would study the entire sequence.

Sac7d aligned to Sac7e (not covered in the '424 patent because it is 79% identical to Seq ID 2).
Sac7d: 1  MVKVKFKYKGEEKEVDTSKIKKVWRVGKMVSFTYDDNGKTGRGAVSEKDAPKELLDMLARAEREK 65  Identity Similarity
          M KV+FKYKGEEKEVDTSKIKKVWRVGKMVSFTYDDNGKTGRGAVSEKDAPKEL+DMLARAE++K       92%      98%
Sac7e: 1  MAKVRFKYKGEEKEVDTSKIKKVWRVGKMVSFTYDDNGKTGRGAVSEKDAPKELMDMLARAEKKK 65

Note: A 49 amino acid core sequence is completely identical.
```

Finding alternative species not covered by the allowed claims of the '424 patent, whether the above recited naturally occurring species or man-made muteins are trivial exercises for one skilled in the art. No reasonable protein chemist looking at this data would doubt that Sac7e could increase the processivity of polymerases if traded out for Sac7d in the constructs in Seq ID No. 9 and SEQ No. ID 10 of the '424 patent.

It is also helpful to take note that three of the references Dr. Wang cited in the patent (Choli et. al. **Exhibit 2**, Baumann et. al. **Exhibit 9**, and McAfee et. al. **Exhibit 10**) contain figures with sequence alignments of Sso7d homologues including Sac7d, Sac7a, and Sac7e. They are repeatedly described as structurally and functionally closely related proteins. The Sac7d construct (figure 2 of the application) was made to support that contention that these homologues would work. Dr. Wang clearly knew about and taught these proteins would work in the invention. No one skilled in the art that reads the patent specification and the referenced papers would have objective reasons to think it wouldn't work.

For these reasons, I submit that a 79% identity to Sso7d using naturally occurring variants is clearly enabled by the specification.

13. ROUTINELY INTRODUCING NON-NATURAL VARIATIONS LOWERS THE PERCENTAGE BELOW 79%.

Using natural variants as a road map a 79% identity is readily available. But man-made modifications can take this 79% identity lower. One can go lower in percent identity by merely combining known deviations from Sso7d. Using the family of Sac7 proteins as a road map one obtains the following hybrid sequence:

Hypothetical 7d: MVKVKVRFKYKGEEKQVDTSKIKKVGRVGKMVSATYDDNGKTGRGAVSEKDAPKELLDMLARAEREKK[2]

The hypothetical protein 7d is 76% identical to Sso7d as shown in the alignment below.

```
Sso7d:     ATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDAPKELLQML---EKQKK 64
           --V+FKYKGEEK+VD-SKIKKV-RVGKM+SFTYD+--GKTGRGAVSEKDAPKELL-ML---E++KK
Hypothet: VKVRFKYKGEEKQVDTSKIKKVGRVGKMVSFTYDD-NGKTGRGAVSEKDAPKELLDMLARAEREKK 65
```

14. COMBINING ALL THE INFORMATION WILL LEAD ONE OF SKILL TO MUTEINS HAVING LESS THAN 60% SEQUENCE IDENTITY TO SAC7d.

Combining all of these changes together one can get a functional derivative of SEQ ID No. 2 with less than 60% amino acid identity in a blast search. One example of such a protein sequence is below.

---

[2] One known Sso7d divergence was not included in this alignment. The F34A mutation was not included because it is known to destabilize the protein. All other divergences are from functional proteins.

VKVRVRFKYK GEERQVDTSR IRKVGRVGKM VSATYDDACA AACNGRTGRG AVSERDAPRE LLDMLARAER ERR

We have identified other muteins of Sso7d that enhance polymerase performance. For instance: I^{21} to V, T^{47} to N, D^{56} to Y, and M^{64} to K in the above sequence. With one exception (I^{21} to V), these are not conserved changes; but, they are changes that do not affect core structures.

When all this information is combined, it would be straightforward to identify muteins with less than 60% identity to Sso7d that would still enhance polymerase performance.

15. THE ARCHAEAL 7 kDa PROTEINS ARE AN ANCIENT PROTEIN AND EXISTING EVOLUTIONARY DRIFT ESTABLISHES THE HIGH PROBABILITY THAT MUTEINS WITH 50% IDENTITY TO ANY KNOWN SPECIES CAN BE CREATED.

From an evolutionary perspective, this family of thermal stable DNA binding proteins is apparently quite ancient. There is a restriction endonuclease from *Methanococcus jannashii* (Results below)--another archaeon-- that a blast search of the Swissprot Database with Seq ID No. 2 will identify. The 47% identity of this DNA binding protein to Sso7d indicates that the DNA binding domain has been around for a long time and that with routine sequencing of genomes from the Archaeal family there will be many easily obtainable proteins with even less than 50% identity to Seq ID No. 2 that will work in the invention.

>gi|10954528|ref|NP_044167.1| M. jannaschii predicted coding region MJECL41 [Methanococcus jannaschii]

gi|12229988|sp|Q60296|T1SH_METJA Putative type I restriction enzyme MjaXP specificity protein (S protein) (S.MjaXP)

gi|2129054|pir||H64514 hypothetical protein MJECL41 - Methanococcus jannaschii plasmid pURB800

gi|1522674|gb|AAC37110.1| M. jannaschii predicted coding region MJECL41 [Methanococcus jannaschii]
Length = 432

Score = 30.0 bits (66), Expect = 8.5
Identities = 19/45 (42%), Positives = 24/45 (53%), Gaps = 1/45 (2%)

```
Query:  3  VKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSE  47
           VKF+++.E.KE.DI.KI.K.W.V.K.I.....+.....GG.T.....+.E
Sbjct:  5  VKFRWETEFKETDIGKIPKDWDV-KKIKDIGEVAGGSTPSTKIKE  48
```

Having provided multiple objective roadmaps to the creation of muteins, it needs to be said that actual function is always subject to empirical determination. To determine if the 7 kDa Archaeal muteins function as desired, the Examiner is asked to take note of the generic assay for DNA binding described on pages 18-19 of the specification. Here, the inventors present a generic method for readily and conveniently testing for operable species.

Based on the objective reasons set forth above, I submit that the creation of Archaeal 7 kDa muteins having 60 to 50% identity to native Archaeal 7 kDa is a matter of routine experimentation.

16. DEFINING THE PROTEINS BY THEIR ABILITY TO BIND TO ANTIBODIES GENERATED AGAINST A PROTOTYPE LIMITS THE PRIMARY AMINO ACID TO DEFINED STRUCTURE.

In addition to defining the invention by a percent identity, an alternative scope of claim protection was presented where the DNA binding proteins were defined as those recognized by polyclonal antibodies generated against specific Archaeal 7 kDa DNA binding proteins. The Examiner has rejected claims directed to non-specific double-stranded nucleic acid binding domains that are recognized by polyclonal antibodies generated against Sso7d. As I understand the rejection, the Examiner believes that the scope of this claim encompasses too many non-operable species to be considered allowable.

In the first instance, I would like to point out that the scope of proteins encompassed by the language is more limited than the claims where the proteins have 50% identity.

The use of immuno-crossreactivity to define proteins as related or unrelated is an old and well-recognized art. The specification, at pages 16-18 provides a routine and conventional means to compare unknown proteins with known proteins.

In addition, it is well-known in the art to use antisera as identification reagents to clone genes, based on the expression of a protein mediated by an expression vector. If the library source is one of the naturally-occurring relatives of *Sulfolobus sulfataricus* listed above, the probability that any cross-reacting gene obtained from the library would function to increase the processivity of polymerases is very high.

But naturally occurring proteins are not the only proteins that would be expected to cross react with polyclonal antisera generated against the prototype Archaeal 7 kDa proteins. One could easily envision muteins that would retain immuno-crossreactivity. To the extent that some may lack function; those inoperable embodiments could be rapidly distinguished from operable species using the prescribed assay set forth in the specification.

When these teachings are coupled with the generic assay for testing functionality of the proteins to non-specifically bind to DNA (see the specification at pages 18 and 19), I submit that there is no objective reason to doubt that the identification of many operable species with 50% or greater sequence identity with SSo7d or Sac7d with polyclonal antibodies specific to the two prototypes would be anything other than routine and expected.

This Declarant has nothing further to say.

Peter B. Vander Horn

Dated: 3/1/04

attachments: Exhibit 1-10

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 415-576-0200
Fax: 415-576-0300
KAW:jhd

60140599 v2

Peter B. Vander Horn, Ph.D.

MJ Bioworks, Inc.
7000 Shoreline Court
South San Francisco, CA 94080
Ph. (650) 635-1324
Fax (650) 635-1302
PeterVH@Bioworks.com

Home:
130 Trimaran Court
Foster City, CA 94404
(650) 358-8776

Expertise: A leader in instrument and reagent development for applications in molecular biology. A comprehensive knowledge in molecular biology, mechanistic enzymology, protein expression and purification, microbiology, nucleotide and fluorescence chemistry. Extensive computer experience.

Current employment at MJ companies

10/2002 to Present	VP of Research, Development, and Engineering, MJ Research & MJ Bioworks
2/2001 to10/2002	VP of Research, MJ Bioworks, Inc.
1/2000 to 2/2001	Acting President & CEO of GeneSys Technologies, Inc.
1997 to 2/2001	Research Director, MJ Bioworks, Inc

Currently managing multiple research and engineering groups totaling about 80 people, including nine Ph.D.s at three facilities (San Francisco, CA; Boston, MA; and Madison, WI). During my tenure, MJ companies have launched several multimillion dollar a year products including real time instruments, thermal cyclers, a DNA sequencer, reagents, and consumables

Responsible for all manufacturing and operations (over 30 non-R&D people) at MJ's Madison, Wisconsin facility. This facility manufactures all of MJ's sequencing and real time instruments, totaling about 25 million dollars in sales in 2003. This facility (formally GeneSys Technologies, Inc.) was acquired in 1999 and fully integrated into the MJ corporate structure in 2001.

Executive sponsor of Oracle ERP implementation.

Some operational responsibilities at MJ's California facility,: supervised laboratory construction, designed 2 high-speed data networks, managed a VoIP implementation, and all HR and IT.

Previous Employment 1993 to 1997	Scientist, Amersham Life Science, Inc. Research Focus: DNA sequencing technologies.
Post-doctoral Experience 1990-1993	Cornell University, Dept. of Chemistry Research Focus: Cloning, overexpression, isolation, and mechanistic enzymology of the thiamine biosynthetic proteins of *Escherichia coli.*
Education	Cornell University

EXHIBIT 1

Ph.D., Microbiology 1990	Dept. of Genetics and Development Minors: Biochemistry and Genetics Research Advisor: Stanley A. Zahler Thesis Research: Regulation of the *ilv-leu* biosynthetic operon of *Bacillus subtilis*.
Summer Research:	Marine Biol. Lab., Woods Hole, MA. 1984. Focus: Microbial Diversity
B.S., Microbiology 1983	Pennsylvania State University University Park, PA.

Experience

Fourteen years experience in the research, development, and launch of many successful commercial products. Four years experience managing a manufacturing facility. Extensive experience in corporate acquisition and integration at both a large and a medium sized company.

Products developed and launched at MJ companies under my leadership include:

- The Opticon™ and Opticon II™ Continuous Fluorescence Detection product lines
- Phusion™ DNA Polymerase Product Line
- Dynamo™ qPCR product line
- The BaseStation™ DNA Sequencer and all software, reagents, and consumables
- New thermal cyclers and updates to older models
- More

Inventor on 3 pending applications and author of several others.

At Amersham I managed the development of reagents for the MegaBACE™ 1000, high throughput capillary-array sequencing instrument, as part of the alliance between Molecular Dynamics and Amersham. While in this role, Molecular Dynamics was acquired and integrated into Amersham.

Managed the development and product launch of:

- DYEnamic Direct™ Fluorescent Sequencing Kits
- DYEnamic Energy Transfer Primers
- *Thermoplasma acidophilum* inorganic pyrophosphatase (for use with Thermo Sequenase)

Assisted in the development and launch of

- Thermo Sequenase
- Thermo Sequenase Radiolabelled Terminator Cycle Sequencing Kit
- TUB™ DNA Sequencing Kit.

Wrote the DYEnamic ET Primer mobility correction files used by ABI 373 and 377 Fluorescent Sequencing Instrument's data analysis program.

Invented the Universal-Shift Energy Transfer Primers; patent pending.

All product launches required a close interaction between manufacturing and commercial groups. I am experienced in the preparation of budgets, business plans, sales support packages, and I have prepared numerous technical articles. I have supervised both technical and marketing trials. I have a thorough knowledge of ISO9000.

I have made multiple presentations at scientific conferences.

Boards of Directors
GeneSys Technologies, Inc 2000 until acquisition
G Corp 1999 to 2003

Taught Human Genetics (2 semesters), Microbial Genetics, and Introductory Microbiology at Cornell University.

Affiliations
American Association for the Advancement of Science
American Society for Microbiology

Personal Interests
Marathon running, cycling, distance swimming.

Publications

Vander Horn, P. B., M. C. Davis, J. J. Cunniff, C. Ruan, B. F. McArdle, S. B. Samols, J. Szasz, G. Hu, K. M. Hujer, S. T. Domke, S. R. Brummet, R. B. Moffett, C. W. Fuller. 1997. Thermo Sequenase DNA Polymerase and *T. acidophilum* Pyrophosphatase: New Thermostable Enzymes for DNA Sequencing. BioTechniques. 22:758-765.

Vander Horn, P. B. and C. W. Fuller. 1996. Fluorescent Energy Transfer for Improved Sensitivity in Fluorescent Dye-Primer DNA Sequencing. Editorial Comments. 24.

Vander Horn, P. B. and C. W. Fuller. 1996. Lab Notes Q&A. Editorial Comments. 24.

C. C. Ruan, C. L. Smith, B. W, Nash, L. P. Hosta, P. B. Vander Horn, and C. W. Fuller. 1996. ^{33}P labeled ddNTPs for DNA Sequencing. Microbial & Comparative: Genomics: Eighth International Genome Sequence and Analysis Conference. Vol.1, No. 3, p250.

Vander Horn, P. B., S. R. Brummet, S. T. Domke, J. J. Holecek, J. A. Krall, C. C. Ruan, and C. W. Fuller. 1996. Fluorescence Energy Transfer Primers and Thermo Sequenase. Microbial & Comparative Genomics: Eighth International Genome Sequence and Analysis Conference. Vol.1, No. 3, p256.

Fuller, C. W., P. B. Vander Horn, J. Cunniff, J. Szasz, R. B. Moffett, and M. C. Davis. 1996. A New Thermostable DNA Polymerase Engineered for DNA Sequencing. Abstracts for the Keystone meeting "DNA Replication and Recombination".

Davis, M. C., P. B. Vander Horn, J. J. Cunniff, J. Szasz, and C. W. Fuller. 1995. New Thermostable Enzymes for DNA Sequencing. Genome Science and Technology: Seventh International Genome Sequence and Analysis Conference. p. 40.

McArdle, B. F., C. Ruan, P. B. Vander Horn, P. Robinson, M. Reeve, and C. W. Fuller. 1995. Optimizing Fluorescent Cycle Sequencing with Thermo Sequenase Polymerase; a New Thermostable Enzyme That Gives Uniform Sequence Band Intensities. Genome Science and Technology: Seventh International Genome Sequence and Analysis Conference. p. 41.

Samols, S. B., P. B. Vander Horn, and C. W. Fuller. 1995. Optimized Radiolabeled Cycle Sequencing with Thermo Sequenase Polymerase; a New Thermostable Enzyme Engineered for Dideoxy Sequencing. Genome Science and Technology: Seventh International Genome Sequence and Analysis Conference. p. 49.

Vander Horn, P. B., and C. W. Fuller. 1996. Fluorescence Energy Transfer for Improved Sensitivity in Fluorescent Dye-Primer DNA Sequencing. Editorial Comments. 23:7.

Vander Horn, P. B., and C. W. Fuller. 1996. Lab Notes: DYEnamic Energy Transfer Primers for Fluorescent DNA Sequencing. In Editorial Comments 23:14-15.

Vander Horn, P. B., E. Mardis, and C. W. Fuller. 1995. Fluorescence Energy Transfer Primers: Brighter Dye Primers for Multicolour DNA Detection and Sequencing. Life Science News. 18:2.

Samols, S. B., B. F. McArdle, C. C. Ruan, P. B. Vander Horn, C. W. Fuller. 1995. Thermo Sequenase; a New Thermostable DNA Polymerase for DNA Sequencing. Editorial Comments. 22:29-36.

Hayle, A. J., A Higgs, M. Raybuck, G. Brophy, D. Parry, and P. B. Vander Horn. 1993. Genome Science and Technology: Fifth International Genome Sequence and Analysis Conference. p. 64.

Vander Horn, P. B., A. D. Backstrom, V. J. Stewart, and T. P. Begley. 1993. Structural Genes for Thiamine Biosynthetic Enzymes (*thiCEFGH*) in *Escherichia coli* K12. J. Bacteriol. 175:982-992.

Vander Horn, P. B., and S. A. Zahler. 1992. Cloning and Nucleotide Sequence of the Leucyl-tRNA Synthetase Gene of *Bacillus subtilis*. J. Bacteriol. 174:3928-3935.

Vander Horn, P. B., A. D. Backstrom, V. J. Stewart, and T. P. Begley. 1992. Cloning, Genetic, and Sequence Analysis of the *thiEFGHJ* Operon of *Escherichia coli*. Abstracts of the 92nd General Meeting of the American Society for Microbiology, p. 279.

Cutting S. M. and P. B. Vander Horn. 1990. Genetic Analysis, p. 27-74. In Colin R. Harwood and Simon M. Cutting (ed.), Molecular Biological Methods for Bacillus. J. Wiley and Sons Limited, West Sussex, England.

Vander Horn, P. B., T. A. Tiernan, A. Keynan, H. Jannasch, and H. O. Halvorson. 1984. A Study of Spore Forming Bacteria Isolated from the Deep Sea. Biological Bulletin, 167:516.

Peter B. Vander Horn, Ph.D.

MJ Bioworks, Inc.
7000 Shoreline Court
South San Francisco, CA 94080
Ph. (650) 635-1324
Fax (650) 635-1302
PeterVH@Bioworks.com

Home:
130 Trimaran Court
Foster City, CA 94404
(650) 358-8776

Expertise: A leader in instrument and reagent development for applications in molecular biology. A comprehensive knowledge in molecular biology, mechanistic enzymology, protein expression and purification, microbiology, nucleotide and fluorescence chemistry. Extensive computer experience.

Current employment at MJ companies

10/2002 to Present	VP of Research, Development, and Engineering, MJ Research & MJ Bioworks
2/2001 to10/2002	VP of Research, MJ Bioworks, Inc.
1/2000 to 2/2001	Acting President & CEO of GeneSys Technologies, Inc.
1997 to 2/2001	Research Director, MJ Bioworks, Inc

Currently managing multiple research and engineering groups totaling about 80 people, including nine Ph.D.s at three facilities (San Francisco, CA; Boston, MA; and Madison, WI). During my tenure, MJ companies have launched several multimillion dollar a year products including real time instruments, thermal cyclers, a DNA sequencer, reagents, and consumables

Responsible for all manufacturing and operations (over 30 non-R&D people) at MJ's Madison, Wisconsin facility. This facility manufactures all of MJ's sequencing and real time instruments, totaling about 25 million dollars in sales in 2003. This facility (formally GeneSys Technologies, Inc.) was acquired in 1999 and fully integrated into the MJ corporate structure in 2001.

Executive sponsor of Oracle ERP implementation.

Some operational responsibilities at MJ's California facility,: supervised laboratory construction, designed 2 high-speed data networks, managed a VoIP implementation, and all HR and IT.

Previous Employment 1993 to 1997	Scientist, Amersham Life Science, Inc. Research Focus: DNA sequencing technologies.
Post-doctoral Experience 1990-1993	Cornell University, Dept. of Chemistry Research Focus: Cloning, overexpression, isolation, and mechanistic enzymology of the thiamine biosynthetic proteins of *Escherichia coli.*
Education	Cornell University

Ph.D., Microbiology 1990	Dept. of Genetics and Development Minors: Biochemistry and Genetics Research Advisor: Stanley A. Zahler Thesis Research: Regulation of the *ilv-leu* biosynthetic operon of *Bacillus subtilis*.
Summer Research:	Marine Biol. Lab., Woods Hole, MA. 1984. Focus: Microbial Diversity
B.S., Microbiology 1983	Pennsylvania State University University Park, PA.

Experience

Fourteen years experience in the research, development, and launch of many successful commercial products. Four years experience managing a manufacturing facility. Extensive experience in corporate acquisition and integration at both a large and a medium sized company.

Products developed and launched at MJ companies under my leadership include:

- The Opticon™ and Opticon II™ Continuous Fluorescence Detection product lines
- Phusion™ DNA Polymerase Product Line
- Dynamo™ qPCR product line
- The BaseStation™ DNA Sequencer and all software, reagents, and consumables
- New thermal cyclers and updates to older models
- More

Inventor on 3 pending applications and author of several others.

At Amersham I managed the development of reagents for the MegaBACE™ 1000, high throughput capillary-array sequencing instrument, as part of the alliance between Molecular Dynamics and Amersham. While in this role, Molecular Dynamics was acquired and integrated into Amersham.

Managed the development and product launch of:

- DYEnamic Direct™ Fluorescent Sequencing Kits
- DYEnamic Energy Transfer Primers
- *Thermoplasma acidophilum* inorganic pyrophosphatase (for use with Thermo Sequenase)

Assisted in the development and launch of

- Thermo Sequenase
- Thermo Sequenase Radiolabelled Terminator Cycle Sequencing Kit
- TUB™ DNA Sequencing Kit.

Wrote the DYEnamic ET Primer mobility correction files used by ABI 373 and 377 Fluorescent Sequencing Instrument's data analysis program.

Invented the Universal-Shift Energy Transfer Primers; patent pending.

All product launches required a close interaction between manufacturing and commercial groups. I am experienced in the preparation of budgets, business plans, sales support packages, and I have prepared numerous technical articles. I have supervised both technical and marketing trials. I have a thorough knowledge of ISO9000.

I have made multiple presentations at scientific conferences.

Boards of Directors
GeneSys Technologies, Inc 2000 until acquisition
G Corp 1999 to 2003

Taught Human Genetics (2 semesters), Microbial Genetics, and Introductory Microbiology at Cornell University.

Affiliations
American Association for the Advancement of Science
American Society for Microbiology

Personal Interests
Marathon running, cycling, distance swimming.

Publications

Vander Horn, P. B., M. C. Davis, J. J. Cunniff, C. Ruan, B. F. McArdle, S. B. Samols, J. Szasz, G. Hu, K. M. Hujer, S. T. Domke, S. R. Brummet, R. B. Moffett, C. W. Fuller. 1997. Thermo Sequenase DNA Polymerase and *T. acidophilum* Pyrophosphatase: New Thermostable Enzymes for DNA Sequencing. BioTechniques. 22:758-765.

Vander Horn, P. B. and C. W. Fuller. 1996. Fluorescent Energy Transfer for Improved Sensitivity in Fluorescent Dye-Primer DNA Sequencing. Editorial Comments. 24.

Vander Horn, P. B. and C. W. Fuller. 1996. Lab Notes Q&A. Editorial Comments. 24.

C. C. Ruan, C. L. Smith, B. W, Nash, L. P. Hosta, P. B. Vander Horn, and C. W. Fuller. 1996. ^{33}P labeled ddNTPs for DNA Sequencing. Microbial & Comparative: Genomics: Eighth International Genome Sequence and Analysis Conference. Vol.1, No. 3, p250.

Vander Horn, P. B., S. R. Brummet, S. T. Domke, J. J. Holecek, J. A. Krall, C. C. Ruan, and C. W. Fuller. 1996. Fluorescence Energy Transfer Primers and Thermo Sequenase. Microbial & Comparative Genomics: Eighth International Genome Sequence and Analysis Conference. Vol.1, No. 3, p256.

Fuller, C. W., P. B. Vander Horn, J. Cunniff, J. Szasz, R. B. Moffett, and M. C. Davis. 1996. A New Thermostable DNA Polymerase Engineered for DNA Sequencing. Abstracts for the Keystone meeting "DNA Replication and Recombination".

Davis, M. C., P. B. Vander Horn, J. J. Cunniff, J. Szasz, and C. W. Fuller. 1995. New Thermostable Enzymes for DNA Sequencing. Genome Science and Technology: Seventh International Genome Sequence and Analysis Conference. p. 40.

McArdle, B. F., C. Ruan, P. B. Vander Horn, P. Robinson, M. Reeve, and C. W. Fuller. 1995. Optimizing Fluorescent Cycle Sequencing with Thermo Sequenase Polymerase; a New Thermostable Enzyme That Gives Uniform Sequence Band Intensities. Genome Science and Technology: Seventh International Genome Sequence and Analysis Conference. p. 41.

Samols, S. B., P. B. Vander Horn, and C. W. Fuller. 1995. Optimized Radiolabeled Cycle Sequencing with Thermo Sequenase Polymerase; a New Thermostable Enzyme Engineered for Dideoxy Sequencing. Genome Science and Technology: Seventh International Genome Sequence and Analysis Conference. p. 49.

Vander Horn, P. B., and C. W. Fuller. 1996. Fluorescence Energy Transfer for Improved Sensitivity in Fluorescent Dye-Primer DNA Sequencing. Editorial Comments. 23:7.

Vander Horn, P. B., and C. W. Fuller. 1996. Lab Notes: DYEnamic Energy Transfer Primers for Fluorescent DNA Sequencing. In Editorial Comments 23:14-15.

Vander Horn, P. B., E. Mardis, and C. W. Fuller. 1995. Fluorescence Energy Transfer Primers: Brighter Dye Primers for Multicolour DNA Detection and Sequencing. Life Science News. 18:2.

Samols, S. B., B. F. McArdle, C. C. Ruan, P. B. Vander Horn, C. W. Fuller. 1995. Thermo Sequenase; a New Thermostable DNA Polymerase for DNA Sequencing. Editorial Comments. 22:29-36.

Hayle, A. J., A Higgs, M. Raybuck, G. Brophy, D. Parry, and P. B. Vander Horn. 1993. Genome Science and Technology: Fifth International Genome Sequence and Analysis Conference. p. 64.

Vander Horn, P. B., A. D. Backstrom, V. J. Stewart, and T. P. Begley. 1993. Structural Genes for Thiamine Biosynthetic Enzymes (*thiCEFGH*) in *Escherichia coli* K12. J. Bacteriol. 175:982-992.

Vander Horn, P. B., and S. A. Zahler. 1992. Cloning and Nucleotide Sequence of the Leucyl-tRNA Synthetase Gene of *Bacillus subtilis*. J. Bacteriol. 174:3928-3935.

Vander Horn, P. B., A. D. Backstrom, V. J. Stewart, and T. P. Begley. 1992. Cloning, Genetic, and Sequence Analysis of the *thiEFGHJ* Operon of *Escherichia coli*. Abstracts of the 92nd General Meeting of the American Society for Microbiology, p. 279.

Cutting S. M. and P. B. Vander Horn. 1990. Genetic Analysis, p. 27-74. In Colin R. Harwood and Simon M. Cutting (ed.), Molecular Biological Methods for Bacillus. J. Wiley and Sons Limited, West Sussex, England.

Vander Horn, P. B., T. A. Tiernan, A. Keynan, H. Jannasch, and H. O. Halvorson. 1984. A Study of Spore Forming Bacteria Isolated from the Deep Sea. Biological Bulletin, 167:516.

Biochimica et Biophysica Acta, 950 (1988) 193–203
Elsevier

BBA 91821

Isolation, characterization and microsequence analysis of a small basic methylated DNA-binding protein from the Archaebacterium, *Sulfolobus solfataricus*

Theodora Choli, Petra Henning, B. Wittmann-Liebold and Richard Reinhardt

Abteilung Wittmann, Max-Planck-Institut für Molekulare Genetik, Berlin (Germany)

(Received 18 December 1987)

Key words: DNA binding protein; Radius of gyration; Amino acid methylation; Microsequence analysis; (*S. solfataricus*)

DNA-binding proteins have been extracted from the thermoacidophilic archaebacterium *Sulfolobus solfataricus* strain P1, grown at 86 °C and pH 4.5. These proteins, which may have a histone-like function, were isolated and purified under standard, non-denaturing conditions, and can be grouped into three molecular mass classes of 7, 8 and 10 kDa. We have purified to homogenity the main 7 kDa protein and determined its DNA-binding affinity by filter binding assays and electron microscopy. The Stokes radius of gyration indicates that the protein occurs as a monomer. The complete amino-acid sequence of this protein contains 14 lysine residues out of 63 amino acids and the calculated M_r is 7149. Five of the lysine residues are partially monomethylated to varying extents and the methylated residues are located exclusively in the N-terminal (positions 4 and 6) and the C-terminal (positions 60, 62 and 63) regions only. The protein is strongly homologous to the 7 kDa proteins of *Sulfolobus acidocaldarius* with the highest homology to protein 7d. Accordingly, the name of this protein from *S. solfataricus* was assigned as DNA-binding protein Sso7d.

Introduction

The mode of packing for eukaryotic DNA is well established. A set of small basic proteins, the histones, are involved in the formation of compact DNA-protein particles which contain the double-helical DNA coiled around an octameric histone complex [1]. In bacteria, the mechanism for folding the long circular DNA molecule into a compact form is much less clear. Although a number of proteins have been implicated for this function [2], a precise description of the composition of 'bacterial chromatin' is not yet available.

Although the structure and composition of the bacterial nucleoids are not very well defined, there is compelling evidence that bacterial DNA is folded into a compact complex [3,4] through the participation of at least three proteins [5]. In recent years, several histone-like DNA-binding proteins have been isolated from eubacteria, called NS1 and NS2, HU, HD or DNA-binding protein II. Their amino-acid sequences have been determined and are currently under further investigation [6–10]. Significant homologies have been found between the eubacterial proteins and the first protein isolated from the archaebacterium

Abbreviations: TPCK, *N*-tosylamido-2-phenylethylchloromethyl ketone; DABITC, 4-*N*,*N*'-dimethylaminoazobenzene-4'-isothiocyanate; SSC, 0.15 M trisodium citrate/0.015 M NaCl (pH 7.0); PMSF, phenylmethylsulphonyl fluoride; BSA, bovine serum albumin; PTH, phenylthiohydantoin.

Correspondence: T. Choli, Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestr. 73, D-1000 Berlin 33 (Dahlem), Germany.

EXHIBIT 2

Thermoplasma acidophilum (for reference see Ref. 8). Previously, at least two groups of DNA-binding proteins with estimated molecular masses of 9 kDa and 6 kDa were found in several *Sulfolobus* species [11]. From our results it has become clear that *Sulfolobus acidocaldarius* contains several DNA-binding proteins of similar sizes with M_r values of 7000, 8000 and 10000 [12,13], of which the predominant protein, 7d [14], and three of the minor components (proteins 7a, 7b and 7e) have been sequenced recently [15].

In this paper we present the isolation, characterization and primary structure determination of the predominant 7 kDa protein from *Sulfolobus solfataricus* strain P1 and compare its sequence with that of the other known bacterial DNA-binding proteins. Our nomenclature for these proteins in the 7 kDa class is based on the increased basicity of the proteins in the order 7a to 7e due to their charge differences [12]. To avoid confusion, it should be pointed out that the primary structure of the dominant 7 kDa protein from *S. acidocaldarius* DSM 1616 has been determined [14], but at those times the organism was named *Sulfolobus solfataricus* DSM 1616. Comparison of DNA-binding proteins, characterization of ribosomal proteins by two-dimensional gel electrophoresis and the immunological characterization of RNA-polymerase subunits had demonstrated clearly that the strain DSM 1616 is similar although not identical to *S. acidocaldarius* DSM 639 and different from other *S. solfataricus* strains [13]. Therefore, this strain was renamed *S. acidocaldarius* DSM 1616.

Experimental procedures

Materials

Sodium dodecylsulfate (SDS) was obtained from Serva (Heidelberg, F.R.G.). TPCK trypsin was obtained from Worthington (Freehold, NJ, U.S.A.). DABITC was from Fluka (Buchs, Switzerland), and recrystallized from boiling acetone. Ovalbumin, chymotrypsinogen A, myoglobin, cytochrome *c* and bovine trypsin inhibitor were from Serva (Heidelberg, F.R.G.). The scintillation cocktail was Beckman Ready-Solv TM[EP], Beckman (Berkeley, CA, U.S.A.). All solutions used for protein purification contained 0.1 mM PMSF, 0.1 mM benzamidine hydrochloride and 6 mM 2-mercaptoethanol, N^{ϵ}-monomethyllysine and the other methylated lysine derivatives were purchased from Serva and CalBiochem (Frankfurt, F.R.G.). Acetonitrile and 2-propanol for HPLC solutions were of LiChrosolv grade and all other chemicals were of pro analysis grade purchased from Merck (Darmstadt, F.R.G.).

Methods

S. solfataricus strain P1 was obtained from W. Zillig (Munich), and cells were grown at 86°C under conditions described in Ref. 12, with the addition of 1 g per liter casamino acids (Difco, Detroit, MI, U.S.A.) to the medium.

Purification of the DNA-binding protein. *S. solfataricus* cells were suspended in Polymix-Hepes buffer [16]. After addition of DNAase I (RNAase free), the cells were broken twice in a Gaulin-Manton press (General Electric, Fort Wayne, IN, U.S.A.) at 72 MPa (9000 lb/inch^2). Cellular debris was removed by centrifugation (1.5 h at $10000 \times g$) and the salt concentration of the supernatant was raised to 1 M NH_4Cl. Ribosomes were separated from smaller proteins by centrifugation overnight at $160000 \times g$. The supernatant was dialysed against 10 mM phosphate buffer at pH 6.0 and applied onto a CM-Sepharose CL-6B column (5 × 40 cm). Proteins were eluted with a linear NaCl gradient from 0.05 to 0.8 M in 10 mM phosphate buffer at pH 6.0 (20 l, flow rate 100 ml/h), 30 ml fractions were collected and assayed for protein content by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Further purification was obtained by gel filtration on Sephadex G-50 superfine in 0.35 M NaCl and additionally by ion-exchange chromatography on Fractogel TSK CM-650 (S) with a linear NaCl gradient from 0.1 to 0.5 M.

Proteins were checked for purity and identified by slab gel electrophoresis in the presence of SDS.

Determination of Stokes radii. Stokes radii of gyration, R_s, were determined by analytical gel filtration on a Sephadex G-50 superfine column (1.7 × 190 cm) in 0.35 M NaCl/20 mM phosphate buffer (pH 7.0). The flow rate was 12 ml/h and the absorption at 230 nm was recorded continuously. The distribution coefficient, k_D, was calcu-

lated from the void volume ((V_0) determined with Dextran blue (2000)), the total available volume ((V_t) determined with benzamidine hydrochloride), and the elution volume (V_e). The calibration line for Stokes radii was obtained by plotting the inverse error function of $(1 - k_D)$ against R_s as described by Ackers [17]. The column was calibrated using the following proteins as markers: ovalbumin (3.0 nm), chymotrypsinogen A (2.2 nm), myoglobin (1.9 nm), cytochrome *c* (1.61 nm) and bovine trypsin inhibitor (1.45 nm).

Filter binding assays. The filter binding assay described in Ref. 18 was modified according to Ref. 13. A fixed amount of ^{3}H-labeled DNA and increasing amounts of protein were incubated in 0.1 × SSC buffer, but containing 0.25 M NaCl, for 15 min at 37°C. DNA-protein complexes were collected onto Millipore filters (0.45 μm, Milford, MA, U.S.A.) which were presoaked for 1 h at 22°C in 10 mM KCl/1 mM EDTA/5 mM 2-mercaptoethanol/50 μg/ml BSA. The complexes were washed three times with 3 ml portions of 0.1 × SSC buffer containing 0.25 M NaCl and quantified by liquid scintillation counting (Beckman LS 7000). The DNA-binding affinity of the examined proteins was expressed in percent referring to the 100% sample of [^{3}H]DNA without protein content.

Gel-filtration binding experiments. DNA binding experiments using size exclusion chromatography on a Sephadex G-50 superfine column (2 × 50 cm) were carried out as described in Ref. 14. A fixed amount of *Sulfolobus* DNA and protein 7d was incubated for 15 min at 67°C in 'polymix' buffer [16]. 1 ml of the sample was injected into the column and comigration of the protein with DNA was established by analysis of the void volume peak by SDS gels.

Electron microscopy studies. The formation of DNA-protein complexes and the preparation of samples for electron microscopy by adsorption to mica was performed as described in Ref. 19. Variable amounts of protein were incubated with double-stranded plasmid RSF 1010 and single-stranded ΦX 174 DNA in a buffer comprising 10 mM triethanolamine-HCl/50 mM KCl/2.5 mM $MgCl_2$/2.5 mM 1,4-dithiothreitol (pH 7.5). Complexes were fixed with 0.2% (v/v) glutaraldehyde, adsorbed to mica and stained with 2% (w/v) aqueous uranyl acetate. Rotary shadowing was done with platinum-iridium (80 : 20) at an angle of about 8°. Electron micrographs were made with a Philips electron microscope, model EM 480.

Enzymatic digestion with trypsin. The protein was digested with TPCK-trypsin (enzyme-to-substrate ratio, 1 : 50) in 100 mM *N*-methylmorpholine acetate buffer at pH 8.1 for 2 h at 37°C, with gentle stirring. The peptides were separated by reversed-phase HPLC (RP-HPLC) on a Vydac C_{18} (201 TPB) column (250 × 4 mm) in dilute aqueous trifluoroacetic acid using an acetonitrile gradient.

Cleavage with CNBr. Protein 7d (1 mg) was cleaved with 6 mg CNBr in 70% (v/v) formic acid for 48 h in the dark under nitrogen at ambient temperature. The peptides obtained were separated directly by RP-HPLC on a Vydac C_4 (214 TP54) column (250 × 4 mm) with a gradient of 2-propanol in aqueous 0.1% trifluoroacetic acid, or with a Vydac C_{18} (201 TPB) column (250 × 4 mm) with an acetonitrile gradient in aqueous trifluoroacetic acid.

Sequence determination. Automatic sequencing of the intact protein was done in a liquid phase sequencer [20] with on-line detection of the PTH-amino acids [21] by isocratic HPLC employing a 2-propanol HPLC solvent system [22] or in a pulsed gas-liquid phase sequencer [23] (Applied Biosystems, model 477A) with on-line detection of the PTH-amino acids by HPLC using a gradient system (Applied Biosystems PTH-analyzer, model 120A). Sequence analysis of tryptic peptides was performed by manual microsequencing employing the DABITC/PITC double coupling method, and the amino-acid derivatives were identified by two-dimensional thin-layer chromatography [24,25]. DABTH-Leu and DABTH-Ile, which comigrate on the micro-TLC plates were identified by isocratic HPLC [26]. The peptides obtained from cyanogen bromide cleavage which carried homoserine residues were sequenced in a solid phase sequencer employing the homoserine lactone attachment procedure [27,28].

Amino-acid analysis. Hydrolysis of the protein and peptides was performed in 100 μl 5.7 M HCl for 24 h at 110°C. The amino acids were determined after precolumn derivatization with *o*-phthaldialdehyde by RP-HPLC separation as described in Ref. 29.

Results and Discussion

The growth of *S. solfataricus* strain P1, breakage of cells and isolation of the DNA-binding proteins were performed as described in the Experimental procedures. Similar to *S. acidocaldarius* cells [12], three molecular weight classes of DNA-binding proteins of 7, 8 and 10 kDa have been isolated from *S. solfataricus* strain P1. The major component of the 7 kDa class is the DNA-binding protein 7d, according to the nomenclature used for the DNA-binding proteins from *S. acidocaldarius* [13].

Fig. 1a shows the protein separation on CM-Sepharose CL-6B. The fractions containing protein 7d and an 8 kDa protein are marked. Further purification of protein 7d was performed by gel-filtration on Sephadex G-50 and by ion-exchange chromatography on CM-Fractogel TSK as described in Experimental procedures (chromatograms not shown). Fig. 1b shows the purified protein 7d from *S. solfataricus* P1 on SDS-PAGE in comparison to 7 kDa DNA-binding proteins from *S. acidocaldarius*.

Stokes radii of gyration

The degree of asymmetry and oligomerisation of proteins are easily determined by analytical gel filtration [17]. This procedure allows the use of low protein concentration in order to avoid artefacts such as protein aggregation. The relation between the Stokes radius, R_s, and the quaternary

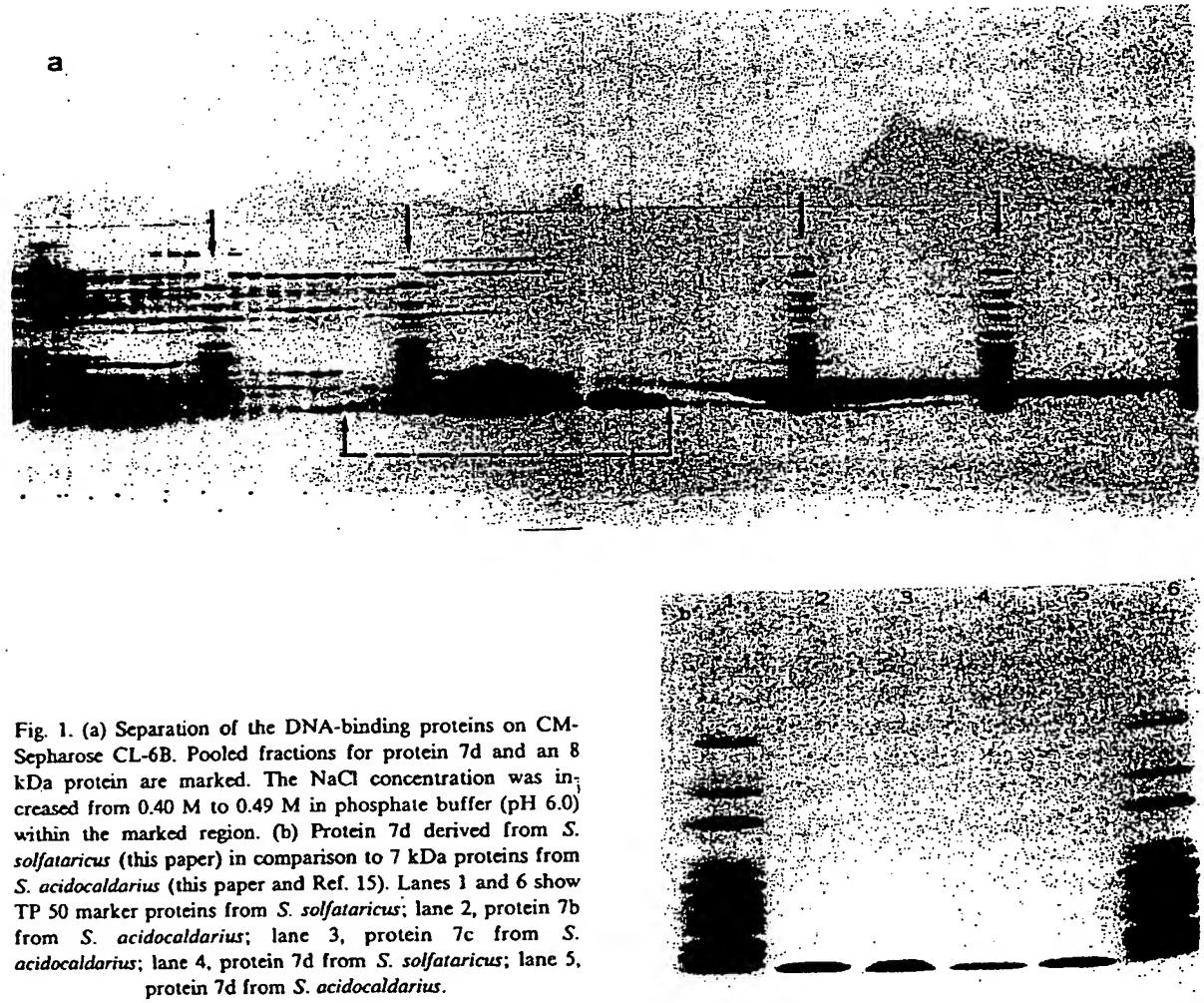


Fig. 1. (a) Separation of the DNA-binding proteins on CM-Sepharose CL-6B. Pooled fractions for protein 7d and an 8 kDa protein are marked. The NaCl concentration was increased from 0.40 M to 0.49 M in phosphate buffer (pH 6.0) within the marked region. (b) Protein 7d derived from *S. solfataricus* (this paper) in comparison to 7 kDa proteins from *S. acidocaldarius* (this paper and Ref. 15). Lanes 1 and 6 show TP 50 marker proteins from *S. solfataricus*; lane 2, protein 7b from *S. acidocaldarius*; lane 3, protein 7c from *S. acidocaldarius*; lane 4, protein 7d from *S. solfataricus*; lane 5, protein 7d from *S. acidocaldarius*.

TABLE I

THE STOKES RADII OF GYRATION OF THE 7 kDa DNA-BINDING PROTEINS FROM *S. ACIDOCALDARIUS* [a] AND *S. SOLFATARICUS* [b] DETERMINED BY ANALYTICAL GEL FILTRATION [17].

The frictional ratio (f/f_0) is calculated from the ratio of R_s and the radius of the equivalent sphere R_{min}.

	R_s (nm)	f/f_0		
		monomer	dimer	tetramer
(a) Sac7d	1.53	1.20	0.95	0.75
(b) Sso7d	1.56	1.21	0.96	0.73

structure of proteins is the frictional ratio, f/f_0, which can be calculated from the experimental R_s value and the theoretical minimal radius, R_{min}, for a given molecular weight. Table I shows that in 0.35 M NaCl the 7 kDa proteins are monomers like the 7 kDa proteins from *S. acidocaldarius*. This is also in accordance with results from ^{1}H-NMR experiments (data not shown).

Filter binding assays

The original procedure [18] for filter binding assays used rather low ionic strength buffer (0.1 × SSC), which allows the nonspecific binding of basic proteins to nucleic acids by electrostatic interactions. In order to avoid this, the NaCl concentration of the binding buffer was increased to 0.25 M in 0.1 × SSC. It has been shown that at this ionic strength, basic proteins like lysozyme, cytochrome *c* or *E. coli* ribosomal proteins do not bind to DNA due to their basicity only [13]. Well established DNA-binding proteins like HU from *E. coli* and DNA-binding protein II from *Bacillus stearothermophilus* showed with these buffer conditions a binding capacity of 18% to 20% at a protein/DNA ratio of 25. The whole set of DNA-binding proteins from *S. acidocaldarius* clearly demonstrated binding capacities in the range of 5% to nearly 80% under the same conditions [12–14]. The filter binding assay of protein 7d (Table II) resulted in a DNA-binding affinity of about 18% binding capacity referring to the 100% sample of [^{3}H]DNA without protein content at a protein/DNA ratio of 25. This value is slightly higher than that of the homologous protein from *S. acidocaldarius*, which can be explained by the different amount of methylated lysines.

The results of the size exclusion experiments confirm qualitatively those from filter binding assays. If the protein/DNA ratio is increased drastically, free protein is fractionated by the Sephadex-G50 superfine column after the void volume peak, which contained the protein/DNA complex. The same results were obtained using either *Sulfolobus* or *E. coli* DNA. In the latter case, incubation temperature was decreased to 37 °C.

Electron microscopy

Fig. 2 presents the electron micrographs of protein 7d in complexed formation with both double- and single-stranded DNA. The formation of the protein-DNA complex results in highly condensed DNA-protein clusters. With increased protein/DNA ratios, the isolated clusters on the DNA merge more and more into a large central protein/DNA cluster, surrounded by loops of free DNA. A preference for single- or double-stranded DNA was not found. Similar structures have been observed for the 7 kDa proteins from *S. acidocaldarius*, which represent a very homogeneous group of five DNA-binding proteins [14,15]. All these highly similar proteins have been shown to interact specifically with single- and double-stranded DNA, although a sequence specificity has not been observed [19].

TABLE II

MILLIPORE FILTER BINDING ASSAYS

Increasing amounts of protein were incubated with 0.5 μg ^{3}H-labeled DNA in the presence of 0.25 M NaCl in 0.1 × SSC. The DNA-binding affinity of protein 7d from *S. solfataricus* is shown. 100% affinity is equivalent to the total amount of [^{3}H]DNA.

Protein/DNA ratio (w/w)	1	5	10	15	20	25
DNA-binding affinity (%)	1	6	10	13	16	18

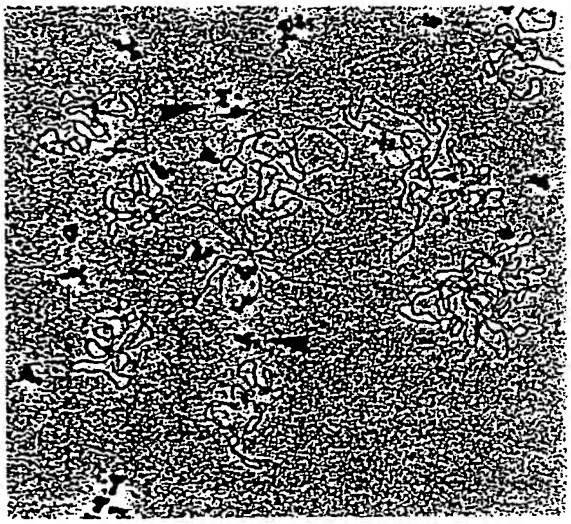


Fig. 2. Electron micrographs of nucleoproteins formed with protein 7d. Some complexes formed with (ss) DNA (ΦX 174) are marked with arrows. Clusters of bound protein on (ds) plasmid DNA RSF 1010, surrounded by free DNA, could be observed.

Amino-acid sequence determination

The complete amino-acid sequence of protein 7d from the archaebacterium *S. solfataricus* and the strategy employed for the sequence determination are shown in Fig. 3. The amino-acid composition derived from the sequence is in good agreement with that obtained from the total hydrolysis of the protein (Table III). As derived from the amino-acid sequence, protein 7d contains modified lysines which were identified as monomethylated residues partially modified at positions 4, 6, 60, 62 and 63 and fully methylated at position 62 (see below).

Occurrence of modified amino acids in the protein

In the PTH-amino acid identification system of the liquid [21,22] and gas-liquid phase sequenator [23], a new peak was observed in steps 4, 6, 60, 62 and 63. This modified derivative was identified on-line as ε-monomethyl-PTH lysine in comparison with an authentic reference.

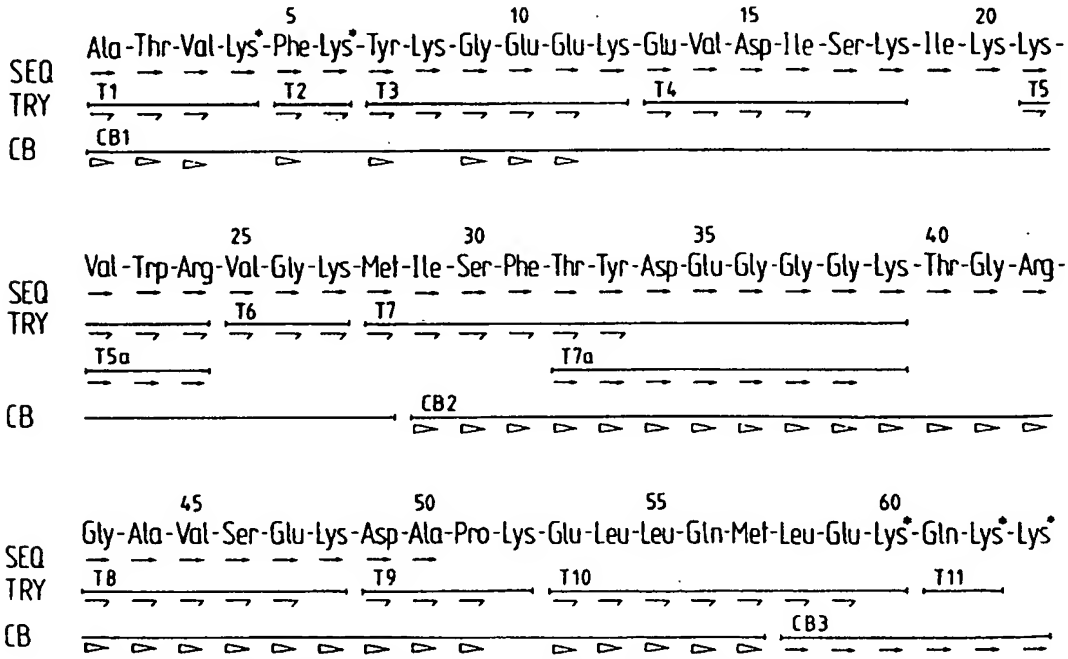
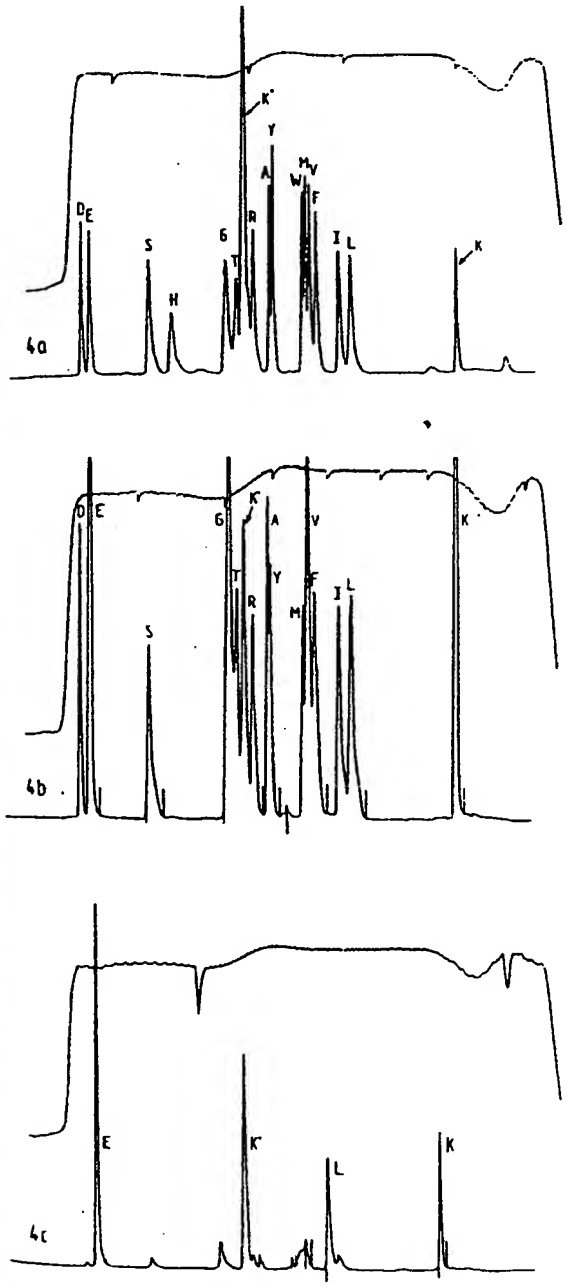


Fig. 3. Amino-acid sequence of DNA-binding protein 7d from *S. solfataricus*. Sequences of individual peptides and intact protein are indicated as follows: →, Sequenced automatically using a pulsed gas-liquid phase sequencer [23], or a liquid-phase sequencer [20–22]. —, Manual liquid-phase DABITC/PITC double coupling method [24,25]. ▷, Solid-phase sequencing after homoserine-lactone attachment to aminopropyl glass (APG) [27,28]. TRY and CB indicate peptides derived from digestion with trypsin or cleavage with CNBr. Lys* indicates the N^{ϵ}-monomethylated lysines.

Furthermore, experiments with lysine derivatives showed that this unusual amino acid comigrates with the authentic *o*-phthaldialdehyde derivative of ε-monomethyl lysine in the amino-acid analyzer [15]. Fig. 4 shows the HPLC separation



of a standard amino-acid mixture plus ε-monomethylated lysine after *o*-phthaldialdehyde derivatization. The additional peak which migrates be-

Fig. 4. (a) Separation of 100 pmol of a reference amino-acid mixture containing N^{ε}-monomethylated lysine, after ortho-phthaldialdehyde precolumn derivatization, by reversed-phase HPLC, using a column (250×4 mm) filled with Shandon Hypersil ODS 5μ material. Buffer A was 12.5 mM Na_2HPO_4 (pH 7.2), and buffer B was 3% tetrahydrofuran in methanol [27]. The peak which appears between threonine and arginine comigrates with authentic ε-monomethylated lysine (K*). (b) The amino-acid composition of protein Sso7d after total hydrolysis. The separation of the amino acids was as described in Fig. 4a. The characteristic peak for N^{ε}-monomethyl lysine (K*) appears at the same position in the chromatogram. (c) The amino-acid composition of the C-terminal peptide (CB 3) after acid hydrolysis. The separation of the amino acids was as described in Fig. 4a. The peak marked with an asterisk shows the ε-monomethylated lysine residue.

TABLE III

AMINO-ACID ANALYSIS OF THE DNA-BINDING PROTEIN 7d FROM *S. SOLFATARICUS*

n.d., residues not determined by amino-acid analysis.

	Number of residues derived by amino-acid:	
	sequence	analysis [a]
Asp	3	2.6
Asn	–	–
Glu	7	9.0
Gln	2	–
Ser	3	2.4
Gly	7	7.6
Thr	3	2.4
Arg	2	2.3
Ala	3	3.0
Tyr	2	1.7
Trp	1	n.d.
Met	2	1.2
Val	5	5.6
Phe	2	1.6
Ile	3	2.9
Leu	3	3.1
Lys [b]	14	12.6
Pro	1	n.d.

[a] The values given are not corrected for destruction of amino acids or incomplete hydrolysis.

[b] Lys refers to the sum of lysine and monomethylated lysine. Due to the presence of incompletely modified lysines, the value for lysines by amino-acid analysis cannot be calculated precisely.

tween threonine and arginine derivatives was determined to be ε-monomethyllysine, whereas ε-dimethyllysine migrated after the arginine derivative and ε-trimethyllysine before glycine.

Fig. 4b shows the separation of the amino-acid derivatives of protein 7d produced after amino-acid hydrolysis. Between the arginine and threonine *o*-phthaldialdehyde derivatives, the ε-monomethyllysine of the hydrolysate of the DNA-binding protein 7d can be identified.

Separation of tryptic peptides and N-terminal sequence region

Fig. 5 demonstrates the separation of the tryptic peptides by RP-HPLC with a Vydac C_{18} column. Some peptides with the same amino-acid composition except for the lysine content elute at different retention times. This effect is probably caused by the different degree of methylation of lysine residues. Sequence information and *o*-phthaldialdehyde-amino-acid determination demonstrates that the peptides $T1_2$ and $T1_4$ have Lys-4 modified, with the sequence Ala-Thr-Val-Lys* (pos. 1–4, see Fig. 3), while peptide $T1_1$ contains an unmodified lysine residues with the sequence Ala-Thr-Val-Lys. Peptide $T1_3$ is a mixture of the peptides $T1_1$ and $T1_2$. Peptide T2, Phe-Lys* (pos. 5–6, see Fig. 3) is found in one position only. The degree of methylation, derived from the sequence of the intact protein and estimated by peak height, is approx. 90% for Lys-4 and 83% for Lys-6.

The appearance of peptide T7 (pos. 28–39), which does not possess modified lysines, at two different positions may be due to partial oxidation of methionine. The degree of modification at Lys-60 appears to be the crucial factor for the elution of peptide T10 (pos. 52–60) at different positions. Amino-acid analysis of this peptide has shown that peptides $T10_1$ and $T10_2$ differ only at Lys-60, namely $T10_1$ contains unmodified lysine, while Lys-60 in $T10_2$ is monomethylated.

C-terminal peptide regions

The peptides produced after CNBr cleavage were separated by RP-HPLC either on a Vydac C_4 or C_{18} column as described in Experimental procedures. The C-terminal peptide (CB3) (pos. 58–63) was isolated by using the Vydac C_{18} column and the homoserine peptides CB1 (pos. 1–28) and CB2 (pos. 29–57) by a Vydac C_4 column. From the sequence determination and amino-acid analysis (Fig. 4c) of CB3, the following primary structure was derived: 58-Leu-Glu-Lys*-Gln-Lys*-Lys*-63. The degree of monomethylation, as estimated by peak height, is approx. 90%, 100% and 58% for lysine residues 60, 62 and 63, respectively. The number of lysine residues in the C-terminal peptide was substantiated by fast atom bombardment mass spectrometry [30].

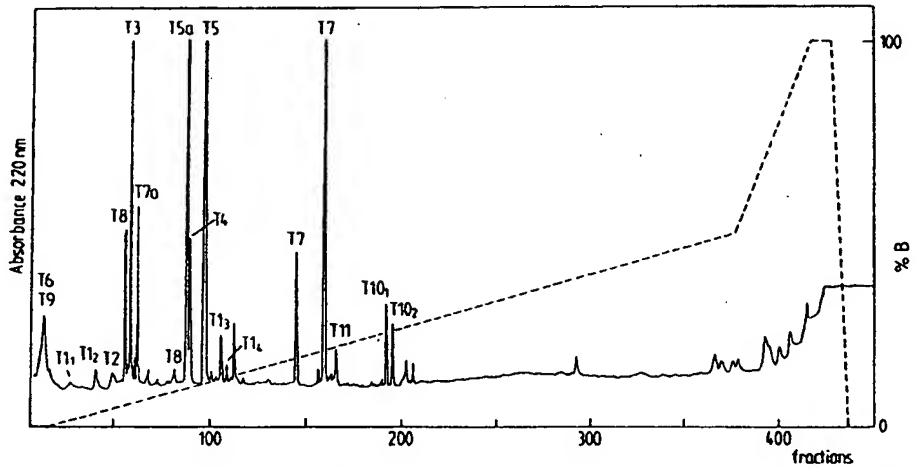


Fig. 5. Separation of the 20 nmol peptides derived by tryptic digestion of protein Sso7d by HPLC. The peptides were chromatographed on a Vydac C_{18} (201 TPB) column (250×4 mm) in a solvent system of 0.1% trifluoroacetic acid/acetonitrile. The gradient applied was 100% A for 10 min, 0–50% B in 180 min, 50–100% B in 20 min, 100% B for 5 min and 100–0% B in 5 min. Measurements were made at 220 nm, 0.16 arbitrary units (full scale), at a flow rate of 1.0 ml/min.

Because of the methylation of the lysines found here in the *S. solfataricus* 7d protein (Sso7d), the homologous 7d protein derived from *S. acidocaldarius* (Sac7d) was also examined for lysine modifications not previously identified [14]. We reinvestigated the Sac7d protein by liquid phase sequencing and isolation of the C-terminal CNBr fragment, and found N^{ϵ}-monomethylated lysines at positions 4 and 6 (approx. 20% and 50%, respectively). However, in contrast to Sso7d, the corresponding Sac7d protein showed no modified lysine residues in the C-terminal sequence region.

Secondary structure predictions

Information about the secondary structure of protein 7d has been predicted based on the amino-acid sequence. Four different prediction methods according to Ref. 31 were used to calculate the conformational states (Fig. 6). This protein possesses a higher amount of α-helical do-

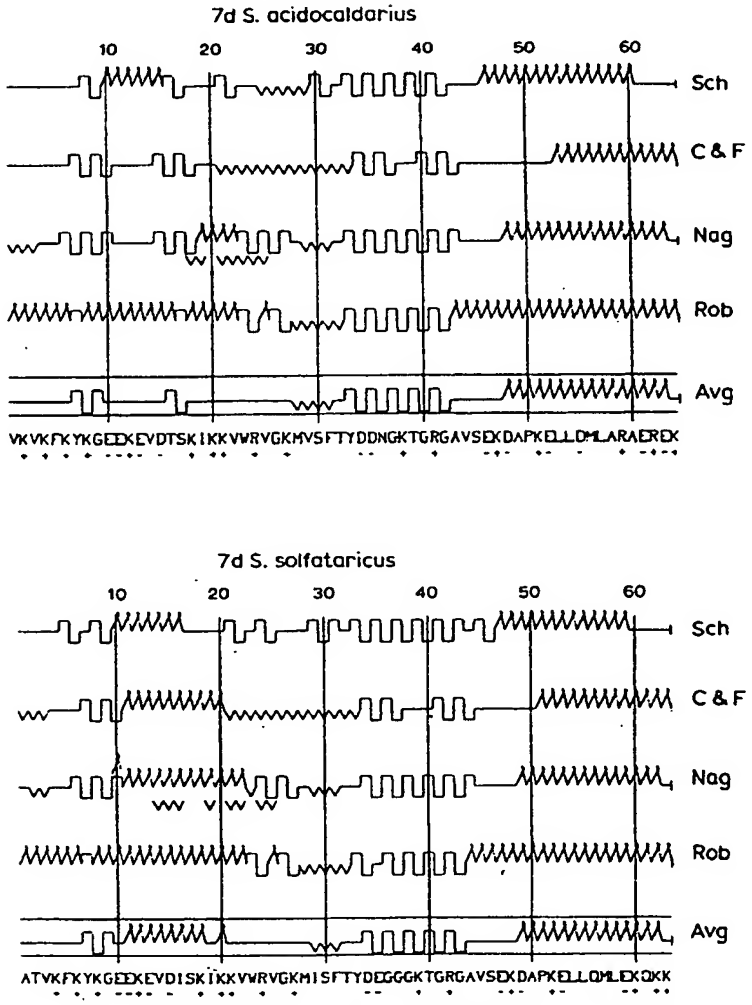


Fig. 6. Secondary structure of DNA-binding proteins 7d from *S. acidocaldarius* and *S. solfataricus* as predicted by four different methods. The symbols represent residues in α-helical (ΛΛΛ), beta-sheet (∿∿∿), β-turns (⊓⊔⊓) and random coil (————) formations. The line Avg summarizes the secondary structure obtained when at least three of the four predictions are in agreement. The amino-acid sequences of the proteins are shown at the bottom line in the one-letter code. Sch, method according to Burgess et al. [33]; C&F, Chou and Fassman [34]; Nag, Nagano [35]; Rob, Robson and Suzuki [36].

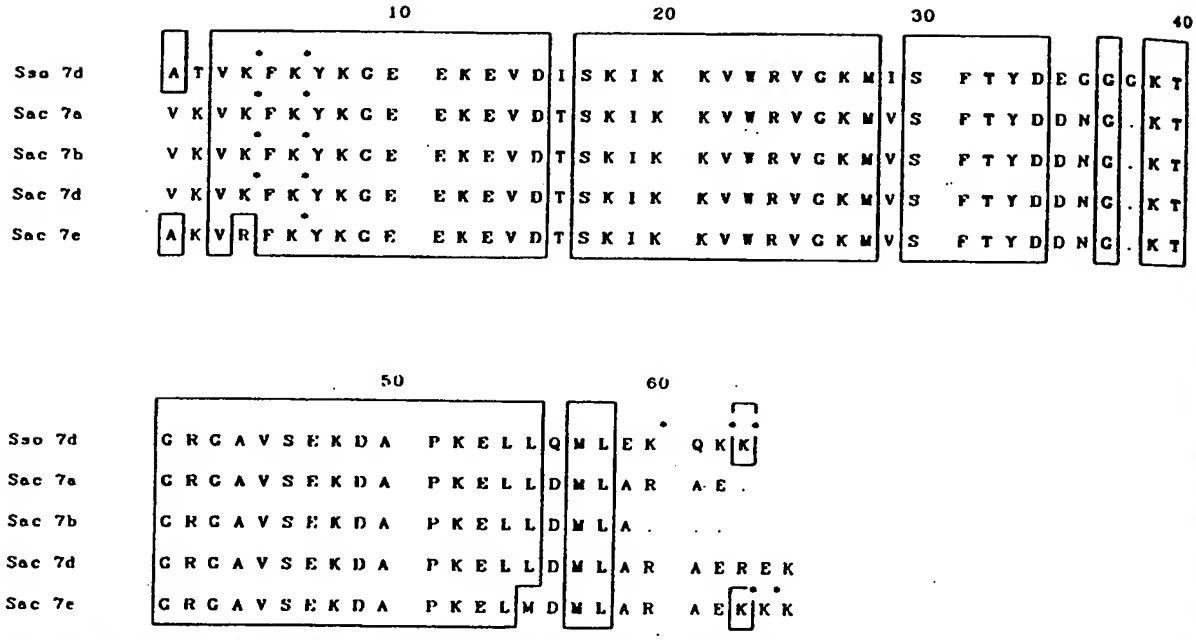


Fig. 7. Structural homology between the 7d DNA-binding protein from *S. solfataricus* and the DNA-binding proteins 7a, 7b, 7e and 7d from *S. acidocaldarius* cells. The alignment scores (SD units) calculated by the program ALIGN [32] using the standard mutation data matrix (100 random runs and a break penalty of 20) are:
7d *S. solfataricus* – 7a *S. acidocaldarius*: 30.93. 7d *S. solfataricus* – 7d *S. acidocaldarius*: 32.63.
7d *S. solfataricus* – 7b *S. acidocaldarius*: 29.54. 7d *S. solfataricus* – 7e *S. acidocaldarius*: 30.23.
Gaps are shown as. . . .

mains – about 35% – as compared to other 7 kDa DNA-binding proteins from *S. acidocaldarius* for which only about 15% helix content was calculated.

Homology to other DNA-binding proteins

By sequence comparison, we found a strong degree of homology between protein 7d from *S. solfataricus* and the proteins from the 7 kDa group from the archaebacterium *S. acidocaldarius* (Fig. 7), using the programme ALIGN [32]. No significant homology between protein 7d from *S. solfataricus* and DNA-binding proteins from other organisms has been found.

Acknowledgements

The authors wish to express their thanks to Dr. H.G. Wittmann for support of the work, Dr. K. Eckart for the mass spectrometry experiments, to Ms. U. Kapp for excellent technical assistance with manual microsequencing, to A. Köpke for help with computing and V. Kruft for typing the paper.

References

1 Baldwin, J.P., Bosely, P.G. and Bradbury, E.M. (1975) Nature 253, 245–249.
2 Geider, H. and Hoffmann-Berling, H. (1981) Annu. Rev. Biochem. 50, 233–260.
3 Worcel, A. and Burgi, E. (1972) J. Mol. Biol. 71, 127–147.
4 Delius, H. and Worcel, A. (1974) J. Mol. Biol. 82, 107–109.
5 Pettijohn, D.E. (1982) Cell 30, 667–669.
6 Mende, L., Timm, B. and Subramanian, A. (1978) FEBS Lett. 96, 395–396.
7 Laine, B., Kmiecik, D., Sautiere, P., Biserte, G. and Cohen-Salal, M. (1980) Eur. J. Biochem. 103, 447–461.
8 Kimura, M. and Wilson, K.S. (1983) J. Biol. Chem. 258, 4007–4011.
9 Hawkins, A.R. and Wrotton, J.C. (1981) FEBS Lett. 130, 275–278.
10 Searcy, D.G. (1975) Biochim. Biophys. Acta 395, 535–547.
11 Thomm, M., Stetter, K.O. and Zillig, W. (1982) Zbl. Bact. Hyg. C3, 128–139.
12 Grote, M., Dijk, J. and Reinhardt, R. (1986) Biochim. Biophys. Acta 873, 405–413.

13 Dijk, J. and Reinhardt, R. (1986) in Bacterial Chromatin (Gualerzi, C.O. and Pon, C.L., eds.), pp. 185–218, Springer-Verlag, Berlin.
14 Kimura, M., Kimura, J., Davie, P., Reinhardt, R. and Dijk, J. (1984) FEBS Lett. 176, 176–178.
15 Choli, T., Wittmann-Liebold, B. and Reinhardt, R. (1988) J. Biol. Chem. 263, in press.
16 Jelence, C.P. (1980) Anal. Biochem. 105, 369–374.
17 Ackers, G. (1976) J. Biol. Chem. 242, 3237–3238.
18 Lammi, M., Paci, M. and Gualerzi, C.O. (1984) FEBS Lett. 170, 99–104.
19 Lurz, R., Grote, M., Dijk, J., Reinhardt, R. and Dobrinski, B. (1986) EMBO J. 5, 3715–3721.
20 Wittmann-Liebold, B. (1983) in Modern Methods in Protein Chemistry (Tschesche, H., ed.), pp. 229–266, W. de Gruyter, Berlin.
21 Wittmann-Liebold, B. and Ashman, K. (1985) in Modern Methods in Protein Chemistry (Tschesche, H., ed.), pp. 303–327, W. de Gruyter, Berlin.
22 Ashman, K. and Wittmann-Liebold, B. (1985) FEBS Lett. 190, 129–132.
23 Hewick, R.H., Hunkapillar, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990–7997.
24 Chang, J.Y., Brauer, D. and Wittmann-Liebold (1978) FEBS Lett. 92, 205–214.
25 Wittmann-Liebold, B., Hirano, H. and Kimura, M. (1986) in Advanced Methods im Protein Microsequence Analysis (Wittmann-Liebold, B., Salnikow, J. and Erdmann, V.A., eds.), pp. 77–90, Springer-Verlag, Berlin.
26 Lehmann, A. and Wittmann-Liebold, B. (1984) FEBS Lett. 176, 360–364.
27 Salnikow, J., Lehmann, A. and Wittmann-Liebold, B. (1981) Anal. Biochem. 117, 433–442.
28 Salnikow, J. (1986) in Advanced Methods in Protein Microsequence Analysis (Wittmann-Liebold, B., Salnikow, J. and Erdmann, V.A., eds.), pp. 108–116, Springer-Verlag, Berlin.
29 Ashman, K. and Bosserhoff, A. (1985) in Modern Methods in Protein Chemistry (Tschesche, H., ed.), pp. 155–171, W. de Gruyter, Berlin.
30 Burlingame, A.L., Baillie, T.A. and Derrick, P.J. (1986) Anal. Chem. 58, 165–168.
31 Rawlings, N., Ashman, K. and Wittmann-Liebold, B. (1983) Int. J. Pep. Prot. Res. 22, 515–524.
32 Dayhoff, M.O. (1978) in Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Washington, DC.
33 Burgess, A.W., Ponnuswamy, P.K. and Scheraga, H.A. (1974) Isr. J. Chem. 12, 239–286.
34 Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–147.
35 Nagano, K. (1977) J. Mol. Biol. 109, 251–274.
36 Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327–356.

1. Murray, A.J., Lewis, S.J., Barclay, A.N. & Brady, R.L. *Proc. Natl. Acad. Sci. USA* **92**, 7337–7341 (1995).
2. Jones, E.Y., Davies, S.J., Williams, A.F., Harlos, K. & Stuart, D.I. *Nature* **360**, 232–239 (1992).
3. Bodian, D.L., Jones, E.Y., Harlos, K., Stuart, D.I. & Davis, S.J. *Structure* **2**, 755–766 (1994)
4. Driscol, P.C., Cyster, G.J., Campbell, I.D. & Williams, A.F. *Nature* **353**, 762–765 (1991).
5. Bennet, M.J., Schlunegger, M.P. & Eisenberg, D., *Prot. Sci.* **4**, 2455–2468 (1995).
6. Peterson, A. & Seed, B. *Nature* **329**, 400–403 (1987).
7. Arulanandam, A.R.N. *et al. Proc. Natl. Acad. Sci. USA* **90**, 11613–11617 (1993).
8. Somoza, C., Driscoll, P.C., Cyster, J.G. & Williams, A.F. *J. Exp. Med.* **178**, 549–558 (1993).
9. van der Merwe, P.A., McNamee, P.N., Davies, S.J. & Barclay, A.N. *EMBO J.* **12**, 4945–4954 (1993).
10. Schlunegger, M.P., Bennet, M.J. & Eisenberg, D. (1997) *Adv. Prot. Chem.* **50**, 61–122 (1997).
11. Barclay, A.N. *et al.* In *The leukocyte antigen factsbook*. pp. 54–62 (Academic Press, New York; 1993).
12. Chothia, C., Gelfand, I. & Kister, A. *J. Mol. Biol.* **278**, 457–479 (1998).
13. Anderson, W.F., Ohlendorf, D.H., Tekeda, Y. & Mattews, B.W. *Nature* **290**, 754–758 (1981).
14. Yan, Y. *et al. Science* **262**, 2027-2030 (1993).
15. Tegoni, M., Ramoni, R., Bignetti, E., Spinelli, S. & Cambillau, C. *Nature Struct. Biol.* **3**, 863–867 (1996).
16. Bianchet, M.A. *et al. Nature Struct. Biol.* **3**, 934–939 (1996).
17. Otwinowski, Z. & Minor, W. *Meth. Enz.*. **276**, 307–326 (1996).
18. Navaza J., *Acta Crystallogr.* A **50**, 157–163 (1994).
19. Brünger, A.T. *X-PLOR version 3.1: A system for X-ray crystallography and NMR* (Yale University Press, New Haven, Connecticut; 1992).
20. Collaborative Computing Project Number 4 - programs for protein crystallography *Acta Crystallogr.* D **50**, 760–763 (1994).
21. Lamzin, V.S. & Wilson, K.S. *Acta Crystallogr.* D **49**, 129–147 (1993).

The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA

Sso7d and Sac7d are two small (~7,000 M_r), but abundant, chromosomal proteins from the hyperthermophilic archaeabacteria *Sulfolobus solfataricus* and *S. acidocaldarius* respectively. These proteins have high thermal, acid and chemical stability. They bind DNA without marked sequence preference and increase the T_m of DNA by ~40 °C. Sso7d in complex with GTAATTAC and GCGT(^{I}U)CGC + GCGAACGC was crystallized in different crystal lattices and the crystal structures were solved at high resolution. Sso7d binds in the minor groove of DNA and causes a single-step sharp kink in DNA (~60°) by the intercalation of the hydrophobic side chains of Val 26 and Met 29. The intercalation sites are different in the two complexes. Observations of this novel DNA binding mode in three independent crystal lattices indicate that it is not a function of crystal packing.

How do sequence-general DNA binding proteins bind to DNA is a fundamental question for understanding genome structure. However, few examples of structures of sequence-general DNA binding proteins bound to DNA are known. The high thermal, acid and chemical stability associated with Sso7d and Sac7d[1] (Fig. 1*a*) makes them an attractive system for structural, thermodynamic and DNA-binding studies[2–5]. Sac7d and Sso7d have unfolding temperatures of greater than 90 °C (at pH 7.5, 0.3 M KCl) and both are acid stable with T_m's of >60 °C at pH 0. The

a

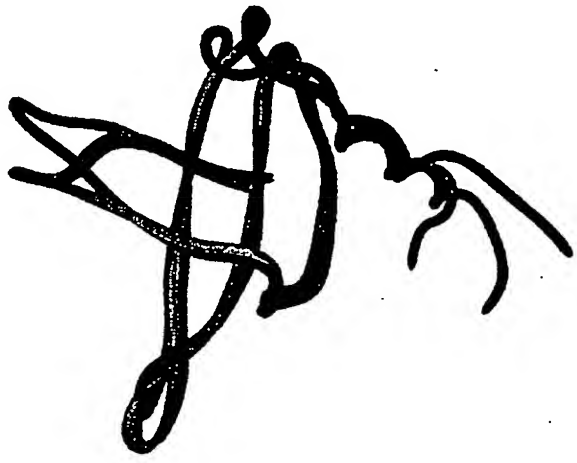
```
                    10                  20                  30
Sso7d   MATVK FKYKG EEKEV DISKI KKVWR VGKMI SFTYD
              —β1—      —β2—      ——β3——|    |——β4——|
Sac7d   -VK-- ----- ----- -T--- ----- ----V -----

                40                  50                  60
Sso7d   EGGGK TGRGA VSEKD APKEL LQMLE KQKK
              ——β5——          ——α-Helix——|
Sac7d   DN-:- ----- ----- ----- -D--A RAERE KK
```

Fig. 1 *a*, Amino acid sequences of recombinant Sso7d and Sac7d. *b*, Ribbon diagram of the Sso7d–GCGT(^{I}U)CGC + GCGAACGC complex. All side chains of Sso7d are shown. The four bridging water molecules are shown as large purple spheres. DNA is colored red for the first two base pairs and green the remaining six base pairs, separated by the intercalating amino acids (yellow). *c*, Superposition of three Sso7d structures from the Sso7d–GCGT(^{I}U)CGC + GCGAACGC complex (yellow), the Sac7d–GCGATCGC complex[6] (green) and the NMR solution structure[6] (red).

b

c



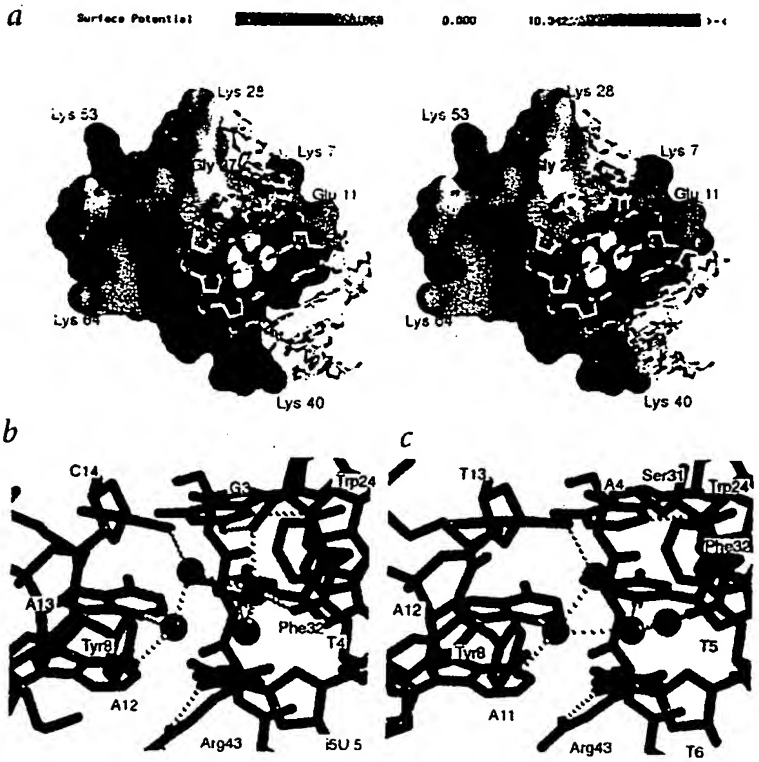


Fig. 2 *a*, Stereoscopic surface drawing of the electrostatic potential of the Sso7d–GCGT(iU)CGC + GCGAACGC complex. The charge distribution of Sac7d was calculated in the absence of DNA. Sso7d is positively charged (+6), resulting from 14 lysines, two arginines, seven glutamic acids and three aspartic acids on the protein surface. However, the complexes are negatively-charged (-8) overall due to the additional 14 negative DNA phosphate charges. There is no apparent correlation between the monomethylation sites of lysines (Lys 5 and Lys 7) and the binding interface. Four bridging waters are found in the space between the protein and DNA. *b*, *c*, Detailed views at the protein–DNA interface of the Sso7d–GCGT(iU)CGC + GCGAACGC (left) and Sso7d–GTAATTAC (right) complexes. Selected side chains of Sso7d (red), three DNA base pairs (green) and four bridging water molecules (purple) are shown.

solution structures of Sso7d[6] and Sac7d[7], determined by NMR analyses, are similar to each other and they consist of an incomplete five-stranded β-barrel capped by an amphiphilic α-helix abutting the β3-β4-β5 strands.

Both proteins bind to DNA without marked sequence preference and increase the T_m of DNA by ~40 °C[3]. However their DNA-binding mode has remained unclear until recently[8]. Baumann *et al.* proposed that the β3-β4-β5 sheet is the putative DNA binding surface[9]. McAfee *et al.*[3] have shown that Sac7d binds to DNA with an average ratio of four base pairs per monomer of Sac7d with no cooperativity. Circular dichroism data also indicated that Sac7d induces a sequence-dependent cooperative structural transition in DNA. Another unusual property is the ribonuclease (RNase) activity associated with Sso7d, which has been called ribonuclease P2[10]. However, similar studies on Sac7d did not produce conclusive evidence of any RNase activity (unpublished results of J.W.S.).

We recently determined the crystal structures of two Sac7d–DNA complexes which revealed an unexpected DNA minor groove binding mode of Sac7d with the DNA duplex sharply kinked[8]. Here we present the results of a parallel study on the structure determination of two Sso7d–DNA complexes. The complexes were crystallized in two new crystal lattices which afford us an excellent opportunity to compare the structure and DNA binding properties of not only the same protein (Sso7d) in different environments, but also different proteins (that is, Sso7d *versus* Sac7d). The structures are also compared with a recent Sso7d–DNA structure by NMR analysis[11].

Overall structure of the complex

The crystal structures of two Sso7d–DNA complexes, Sso7d–GCGT(iU)CGC + GCGAACGC (iU, 5-iodo-deoxyuridine) and Sso7d–GTAATTAC have been solved and well-refined at high resolution (Table 1). All φ/ψ angles of the Ramachandran plot and other conformational parameters in both complexes fall within the acceptable regions. The Sso7d binding sites in DNA are sharply kinked and located at different places in the two complexes: at the C2pG3 step in the Sso7d–GCGT(iU)CGC + GCGAACGC complex (Fig. 1*b*) and at the A3pA4 step in the Sso7d–GTAATTAC complex respectively. The protein covers four bases and significantly widens the DNA minor groove. The other end of the DNA duplex remains B-DNA-like. These two complexes have different crystal packing interactions, indicating that the observed novel DNA binding mode is not a result of crystal packing and is an accurate reflection of the preferred protein–DNA interaction.

The structures of the bound Sso7d in both complexes are very similar to each other with an r.m.s.d. of 0.51 Å (using Cα atoms of residues 2–60) and are generally similar to that of the free Sso7d determined by 2D-NMR analysis[6] with an r.m.s.d. of ~2.10 Å (using Cα atoms of residues 2–60). Some differences exist in the orientation of the β1-β2 β-hairpin and in the conformations of the C-terminal α-helix (Fig. 1*c*).

The molecular surface of Sso7d is irregular with numerous ridges and valleys (Fig. 2*a*). The excellent matching of shapes and charges between Sso7d and DNA in the complexes is evident. A long groove is visible which is occupied by DNA in the complex. There is also a significant crater created by the crossing of the β3-β4-β5 triple stranded β-sheet and the β1-β2 β-hairpin.

Sso7d has a OB-fold topology[12], with a small hydrophobic core of only 11 residues (<25% solvent accessibility). Four glycines (Gly 10, 27, 38 and 39) are located in the loop regions. Many hydrophobic amino acids are solvent exposed (>45% solvent accessibility). The surface hydrophobic amino acids Trp 24, Val 26, Met 29, and Ala 45 are involved in DNA binding contacts.

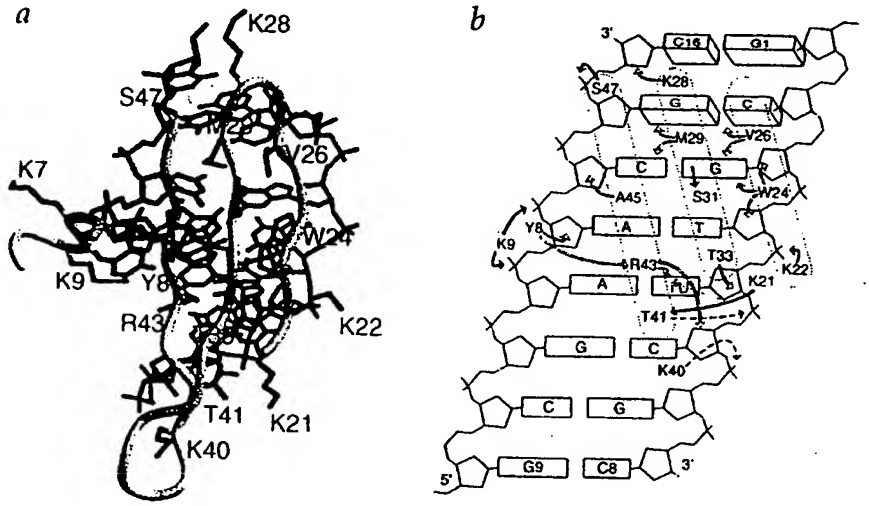


Fig. 3 *a*, Detailed local structures at the protein–DNA interface of the Sso7d– GCGT(ᴵU)CGC + GCGAACGC complex. Selected side chains of Sso7d are shown. *b*, Schematic diagram summarizing all the important Sso7d–DNA contacts. The filled, open and dashed arrows represent direct hydrogen bonds/salt bridges, van der Waals close contacts, and potential hydrogen bonds/salt bridges respectively.

There are two 3_{10}-turns that allow the protein's main chain to change direction abruptly. The C-terminal helix is solvent exposed. A notable feature is the triple-stranded β-sheet (β3-β4-β5) whose interactions with DNA are summarized in Fig. 3.

Bound DNA has a sharp kink

The DNA is severely kinked (Fig. 4) by the bound Sso7d, as in the Sac7d–DNA structures[8]. This type of DNA kink has been observed in the complexes of TBP[13,14] and LEF-1[15], SRY[16] (two HMG-box containing proteins) with their cognate specific DNA sequences, but different from that from proteins that bend DNA more smoothly[17]. The induced local DNA deformation is similar among different protein–DNA complexes, despite the different protein motifs. It should be noted that the ~61° single step kink in the Sso7d–DNA and Sac7d–DNA complexes is the largest among all known structures of protein–DNA complexes. The solution structure of the Sso7d–CTAGCGCGCTAG complex has been analyzed NMR recently[11] and the DNA was found to be bent by 30°, significantly lower than that found in the crystal structures. The difference may be the result of the NMR refinement using limited number of observed NOE crosspeaks between Sso7d and DNA due to the fast exchange between the free and bound DNA/protein.

The bound DNA has a varying degree of helix unwinding at steps surrounding the intercalation sites (-14° at C2pG3, -14° at G3pA4 and -12° at T4piU5). There is also a slight roll (11°) between the G3–C14 and A4–T13 base pairs, thus creating a total bend of 72°. Many nucleotides surrounding the wedge site adopt the less-common C3'-endo (N-type) sugar puckers: C2 (N), G3 (S), T4 (N) and ᴵU5 (N) in one strand and G15 (S), C14 (N), A13 (N) and A12 (S) in the other strand. The Sso7d–GTAATTAC complex has the same pattern.

The DNA distortion seen in the complex described here most likely represents the structural transition observed by McAfee *et al.*[3] using CD spectroscopy for the Sac7d system and the large heat capacity change upon DNA binding observed by Lundback *et al.*[4]. Such a structural transition (unwinding and/or bending) is induced in DNA by Sac7d which is 'cooperative' in the sense that it is necessary to have two proteins bound within a specified distance (for example, 5 base pairs in duplex poly(dA–dT)) before the transition occurs. The inherent resistance to the transition is apparently negligible in short DNA sequences. Our preliminary 1D-NMR titration of Sac7d to cisplatin-lesioned DNA indicates a tight binding between Sac7d and the pre-kinked DNA, supporting the novel binding mode observed in the crystal structures (unpublished data).

Protein–DNA interface

The binding of Sso7d to the minor groove of DNA involves a large

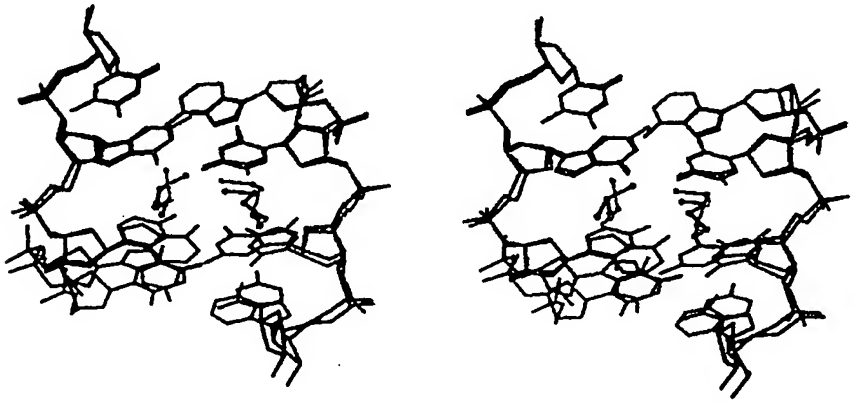


Fig. 4 Stereoscopic view of the intercalation sites. The local structures of the two Sso–DNA complexes are superimposed. The DNA octamer is kinked 61° at the C2pG3 step in the Sso7d–GCGT(ᴵU)CGC + GCGAACGC complex and 62° at the A3–A4 step in the Sso7d–GTAATTAC complex. The sharp kink is due to the intercalation of Val 26 and Met 29 amino acid side chains into DNA base pairs from the minor groove direction, widening the minor groove at this step. The insertions of β4-Met 29 and β3-Val 26 amino acid side chains are ~1.5 Å deep. The side chain of Met 29 lies close to the base pair with the S-CH_3 moiety wedged between the C14 and G15 bases. Similarly the side chain of Val 26 is wedged between the C2 and G3 bases, with each of the δCH_3 groups pointing toward a base.

binding surface area of about 20 Å × 20 Å (Fig. 2*a*). A significant component of the free energy of binding is due to non-electrostatic interactions, made in large part by Trp 24, Val 26 and Met 29 (Fig. 4). In addition to the obvious role of the β4-Met 29 and β3-Val 26 amino acids, the single tryptophan (β3-Trp 24) plays multiple roles. First its bulky ring fills up the space between DNA and Sso7d. Second its indole NH group forms a specific hydrogen bond (2.93 Å) to the N3 of the G3 base, anchoring G3 in its open (unstacked) position. Ala 45 also makes a close van der Waals contact with the deoxyribose of C14, suggesting the requirement of a small hydrophobic side chain of alanine at position 45. Ser 31 receives a hydrogen bond (2.87 Å) from the N2 amino group of the G3 nucleotide. Interestingly, in the Sso7d–GTAATTAC complex Ser 31 forms a hydrogen bond to the sulfur atom of Met 29.

The guanidinium group of Arg 43 is hydrogen bonded to the O2 atom of ^{i}U5. Arg 43 is held in its place with the aid of Tyr 8 whose aromatic ring is stacked on the deoxyribose of A13. The phenolic OH group of Tyr 8 is linked to the N3 of A13 through a bridging water. The hydrogen bond between Arg 43 and the O2 atom of a thymine appears to be important and may determine the polarity of the Sso7d binding mode. The structure of the Sso7d–GCGT(^{i}U)CGC + GCGAACGC complex shows that the Arg 43 of Sso7d is hydrogen bonded to ^{i}U5 of the TT-strand, not to the AA strand. Therefore a combination of the specific interaction between a guanine base and Ser 31, and the hydrogen bond between Arg 43 and ^{i}U5-O2, may be important in favoring the intercalation site at the C2pG3 step in this complex.

The formation of the complex is accomplished by specific hydrogen bonds/salt bridges (Fig. 3). The number of salt bridges between the protein and DNA is in excellent agreement with the five ionic interactions predicted by the salt back-titrations of the Sac7d complex[3] using the theory of de Haseth *et al.*[18]. A somewhat smaller value has been determined by salt-dependent isothermal titration calorimetry on the binding between Sso7d and poly(dG–dC)[4].

An important question is how do Sso7d and Sac7d bind to DNA in a sequence-general manner. The answer may lie in the bridging water molecules found in the buried cavity located between protein and DNA (Fig. 2*b*,*c*). This cavity permits the G–C base pairs to be bound without steric clash due to the additional N2 amino groups, thus endowing the protein with a property required for its sequence-general binding to DNA. For example, in the Sso7d–GCGGTCGC + GCGACCGC complex (which has a G–C base pair, instead of an A–T base pair, at the fourth position in the sequence), we observed fewer intervening water molecules with a concomitant movement of DNA base pairs (unpublished results). It is interesting to note that bridging water molecules play an important role in modulating the sequence-general binding of Sac7d and Sso7d by acting as filler, whereas they play an entirely different role as specific linkers between protein and DNA in defining the sequence specificity in the *Trp* repressor–DNA recognition[19].

Table 1 Crystal and refinement data of two Sso7d-DNA complexes

	Sso7d + GTAATTAC	I-dU-02	I-dU-06	Sso7d + GCGT(^{i}U)CGC +GCGAACGC
Crystal data				
a (Å)	47.60	47.52	47.78	46.87
b (Å)	50.77	50.76	50.91	49.67
c (Å)	42.06	41.97	42.03	37.65
Resolution (Å)	2.0	2.0	2.0	1.7
# reflections (>1.0 σ(F))	7,607	7,499	7,669	11,959
Temperature (°C)	20	20	20	-150
R_{merge} (%)	7.53	6.37	7.33	7.37
Completeness (%)	94.1	92.9	95.7	84.3
Completeness at highest shell for >2.0 σ(F) (%)	83.0 (2.0–2.1 Å)			90.6 (1.70–1.78 Å)
Wilson B-factor (Å^2)	32.6	29.7	32.1	17.8
Mean overall figure of merit	0.83			
Refinement data				
# reflection (>2.0 σ(I))	5,682			9,488
$R_{working}/R_{free}$ (10% data)	0.168/0.268			0.203/0.283
R.m.s.d. bond distance (Å) (Sso7d/DNA)	0.010/0.007			0.014/0.009
R.m.s.d. bond angle (°) (Sso7d/DNA)	1.37/1.20			1.81/1.34
No. of atoms (Sso7d/DNA)	510/322			502/322
No. of waters	99			114

Biological implication

The structures of the Sso7d–DNA and Sac7d–DNA complexes offer new insights into the possible role of several classes of DNA binding proteins in transcription regulation. Some of those proteins, including TBP[13,14], SRY[15], LEF1[16] and PurR[20], bind in the minor groove of DNA and kink the DNA duplex to a different degree[17]. Additionally we noted a possible structural alignment between Sso7d/Sac7d and the cold shock proteins CspA/CspB[21,22]. Both CspA and CspB are related to a class of proteins called Y-box proteins, which have a wide-spread and highly conserved nucleic acid-binding motif occurring from bacteria to humans[23]. Therefore this structural alignment between Sso7d/Sac7d and CspA/CspB may be significant in understanding the Y-box proteins.

The new DNA binding mode of Sso7d/Sac7d may also offer a clue for understanding the packaging of DNA in archaeabacteria. Several models of the polymeric Sso7d–DNA complex with different protein/DNA ratios can be constructed by using the structure of the complex observed in the crystals. Previously we presented a model in which the DNA is maximally-loaded with Sac7d proteins[8]. Additional modeling studies showed that if the number of base pairs per protein monomer is increased (for example, to 10 base pairs per protein), many possibilities for DNA condensation may exist (data not shown)

Our study augments the understanding of chromatin structure in achaea. On the one hand, histones[24] or histone-like proteins (for example, HMf)[25] form nucleosomes. On the other hand, Sso7d/Sac7d may bind to DNA in the minor groove and form higher ordered structures. Thus two different types of DNA compaction mechanisms may be possible in the Archaea: the

mechanism described here with Sac7d/Sso7d which may be representative of the Crenarchaeota, and a nucleosome-like structure for the HMf-class of proteins found in the Euryarchaeota[26,27].

Interestingly, the bacterial HU protein has a different way of forming chomatin structure. The crystal structure of the complex between the integration host factor (IHF) and DNA revealed that IHF induces two prominent kinks in the bound DNA, forming a U-turn[28], by the partial intercalation of a proline from each of the two long β-hairpins which wrap around the DNA. The sequence and structural homology between IHF and HU suggest that HU may organize chromatin using a minor-groove binding mode through intercalation.

Methods

The purified Sso7d protein[2] was dialyzed against de-ionized water and lyophilized. The complexes were crystallized using the vapor diffusion method. The Sso7d–GTAATTAC complex and the two iodo derivatives were crystallized from 1.3 mM Sso7d, 1.3 mM DNA duplex, 2 mM Tris Cl buffer (pH 6.5), 2.6% PEG400 solution, equilibrated with 15% PEG400. The Sso7d–(GCGTTCGC + GCGAACGC) and iodo complexes were crystallized under similar conditions except 5% 2-methyl-2,4-pentanediol (MPD) solution was used and the solution equilibrated with 20% MPD. Data were collected either at room temperature (20 °C) or at -150 °C on a Rigaku R-Axis IIc image plate area detector system to various resolution ranges (Table 1). The crystals of both complexes are in the space group $P2_12_12_1$. The data were processed using the software provided by Molecular Structure Corporation.

The phases were determined by the multiple isomorphous replacement (MIR) method using two iodo derivatives (denoted as I-dU-02 and I-dU-05 with the iodine located at positions T2 and T5 respectively) for the Sso7d–GTAATTAC complex. The figure-of-merit weighted MIR map with solvent flattening at 2.5 Å resolution clearly revealed both the DNA and the Sso7d protein electron density. At that point the refined structure of the Sac7d–GTAATTAC complex[8] was used to model the Sso7d–GTAATTAC complex into the MIR electron density. The model was appropriately corrected against the un-biased map. The structure was refined by the simulated annealing (SA) procedure incorporated in X-PLOR[29] using the data with $|F_o| > 4\sigma(F)$ in the 6.0–1.9 Å range. Simulated annealing and individual temperature factor refinements were carried out by X-PLOR. Well-ordered water molecules were located and gradually included in the model.

Crystals of the complex between Sso7d and GCGTTCGC + GCGAACGC, and the iodo-dU derivatives were obtained. It was found that the sequence GCGT(iU)CGC + GCGAACGC produced the best crystals and a 1.6 Å resolution data set was collected at -150 °C. The structure of the complex was solved by the molecular replacement method using the AMORE package in the CCP4 suite[30]. Similar SA refinement was carried out with a final R-factor (working set) of 20.3% using the $|F_o| > 4\sigma(F)$ data in the 6.0–1.6 Å range.

Programs O[31], MIDAS Plus (University of California at San Francisco) and QUANTA (version 4.0, Molecular Simulation, Burlington, MA) were used to examine the electron density maps and molecular models. The electrostatic potential diagram was calculated by GRASP[32]. DNA force field parameters of Parkinson *et al.*[33] were used. All structures have been refined by SA and individual B-factor refinement in X-PLOR. During the refinement, some rebuilding of the model was necessary to improve the fitting of the model to the electron density. The crystal data and refinement summaries are listed in Table 1.

Coordinates. The atomic coordinates of the two Sso7d–DNA complexes have been deposited in the Brookhaven Protein Data Bank (accession numbers 1BNZ and 1BF4 respectively).

Acknowledgment

This work was supported by the NIH (A.H.J.W. and J.W.S.).

Yi-Gui Gao[1], Shao-Yu Su[1], Howard Robinson[1], Savita Padmanabhan[2], Louis Lim[2], Bradford S. McCrary[2], Stephen P. Edmondson[2], John W. Shriver[2] and Andrew H.-J. Wang[1]

[1]Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [2]Department of Medical Biochemistry, School of Medicine, Southern Illinois University, Carbondale, Illinois 62901, USA.

Correspondence should be addressed to A.H.-J.W. *email: ahjwang@uiuc.edu*

Received 28, May, 1998; accepted 29, July, 1998.

1. McAfee, J. G., Edmondson, S. P., Datta, P. K., Shriver, J. W. & Gupta, R. *Biochemistry* **34**, 10063–10077 (1995).
2. McCrary, B. S., Edmondson, S. P. & Shriver, J. W. *J. Mol. Biol.* **264**, 784–805 (1996).
3. McAfee, J. G., Edmondson, S. P., Zegar, I. & Shriver, J. W. *Biochemistry* **35**, 4034–4045 (1996).
4. Lundback, T., Hansson, H., Knapp, S., Ladenstein, R. & Hard, T. *J. Mol. Biol.* **276**, 775–784 (1998).
5. Dijk, J. & Reinhardt, R. in *Bacterial Chromatin* (eds Gualerzi, C. O. & Pon, C. L.), 185–218 (Springer-Verlag, Berlin; 1986).
6. Baumann, H., Knapp, S., Lundback, T., Ladenstein, R. & Hard, T. *Nature Struct. Biol.* **1**, 808–819 (1994).
7. Edmondson, S. P., Qiu, L. & Shriver, J. W. *Biochemistry* **34**, 13289–13304 (1995).
8. Robinson, H. *et al. Nature* **392**, 202–205 (1998).
9. Baumann, H., Knapp, S., Sarshikoff, A., Ladenstein, R. & Hard, T. *J. Mol. Biol.* **247**, 840–846 (1995).
10. Fusi, P. *et al. Eur. J. Biochem.* **211**, 305–310 (1993).
11. Agback, P., Baumann, H., Knapp, S., Ladenstein, R. & Hard, T. *Nature Struct. Biol.*, **5**, 579–584 (1998).
12. Murzin, A. G. *EMBO J.* **12**, 861–867 (1993).
13. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. *Nature* **365**, 512–520 (1993).
14. Kim, J. L., Nikolov, D. B. & Burley, S. K. *Nature* **365**, 520–527 (1993).
15. Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. *Cell* **81**, 705–714 (1995).
16. Love, J. J. *et al. Nature* **376**, 791–795 (1995).
17. Dickerson, R. E. *Nucleic Acids Res.* **26**, 1906–1926 (1998).
18. de Haseth, P., Lohman, T. & Record, M. T. Jr. *Biochemistry* **16**, 4783–4790 (1977).
19. Otwinowski, Z. *et al. Nature* **335**, 321–329 (1988).
20. Shumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. *Science* **266**, 763–770 (1994).
21. Schindelin, H., Marahiel, M. & Heinemann, U. *Nature* **364**, 164–168 (1993).
22. Schindelin, H., Jiang, W., Inouye, M. & Heinemann, U. *Proc. Natl. Acad. Sci. USA* **91**, 5119–5123 (1994).
23. Wolffe, A. P. *BioEssays* **16**, 245–251 (1994).
24. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. *Nature* **389**, 251–260 (1997).
25. Starich, M. R., Sandman, K., Reeve, J. N. & Summers, M. F. *J. Mol. Biol.* **255**, 187–203 (1996).
26. Grayling, R., Sandman, K. & Reeve, J. *FEMS Microbiol. Rev.* **18**, 203–213 (1996).
27. Ronimus, R. & Musgrave, D. R. *FEMS Microbiol. Rev.* **134**, 79–84 (1995).
28. Rice, P. A., Yang, S.W., Mizuuchi, K. & Nash, H. *Cell* **87**, 1295–1306 (1996).
29. Brünger, A. T. *X-PLOR 3.1, A System for X-ray Crystallography and NMR*. New Haven, Connecticut: Yale University Press, 1992.
30. Collaborative Computational Project, Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr.* **D50**, 760–763 (1994).
31. Jones, T. A., Zou, J. Y., Cowna, S. W. & Kjeldgaard, M. *Acta Crystallogr.* **A47**, 110–119 (1991).
32. Nicholls, A., Sharp, K. & Hoing, B. *GRASP Manual.* (Columbia University, New York, New York, 1992).
33. Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T. & Berman, H. M. *Acta Crystallogr.* **52**, 57–64 (1996).

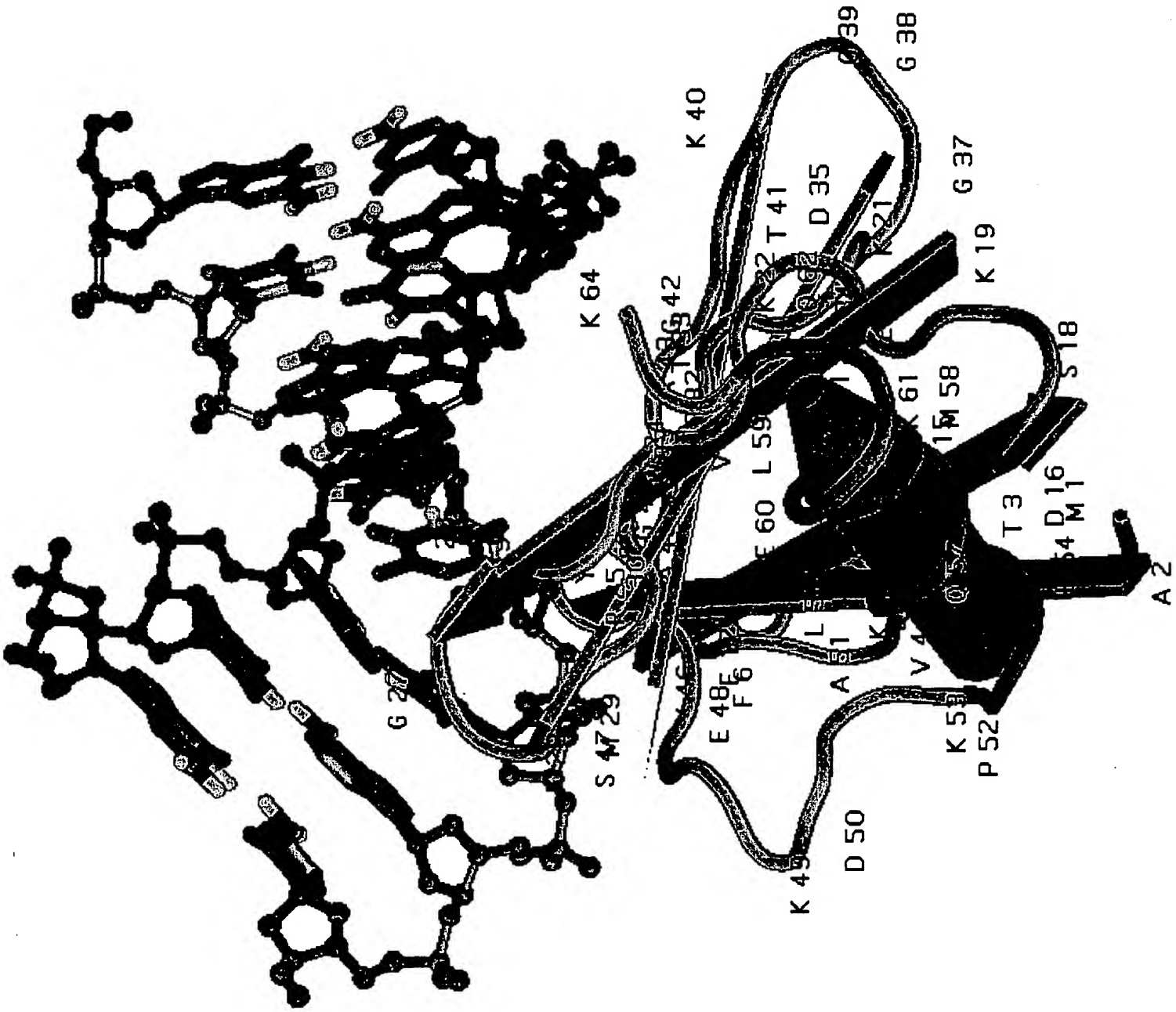


EXHIBIT 4

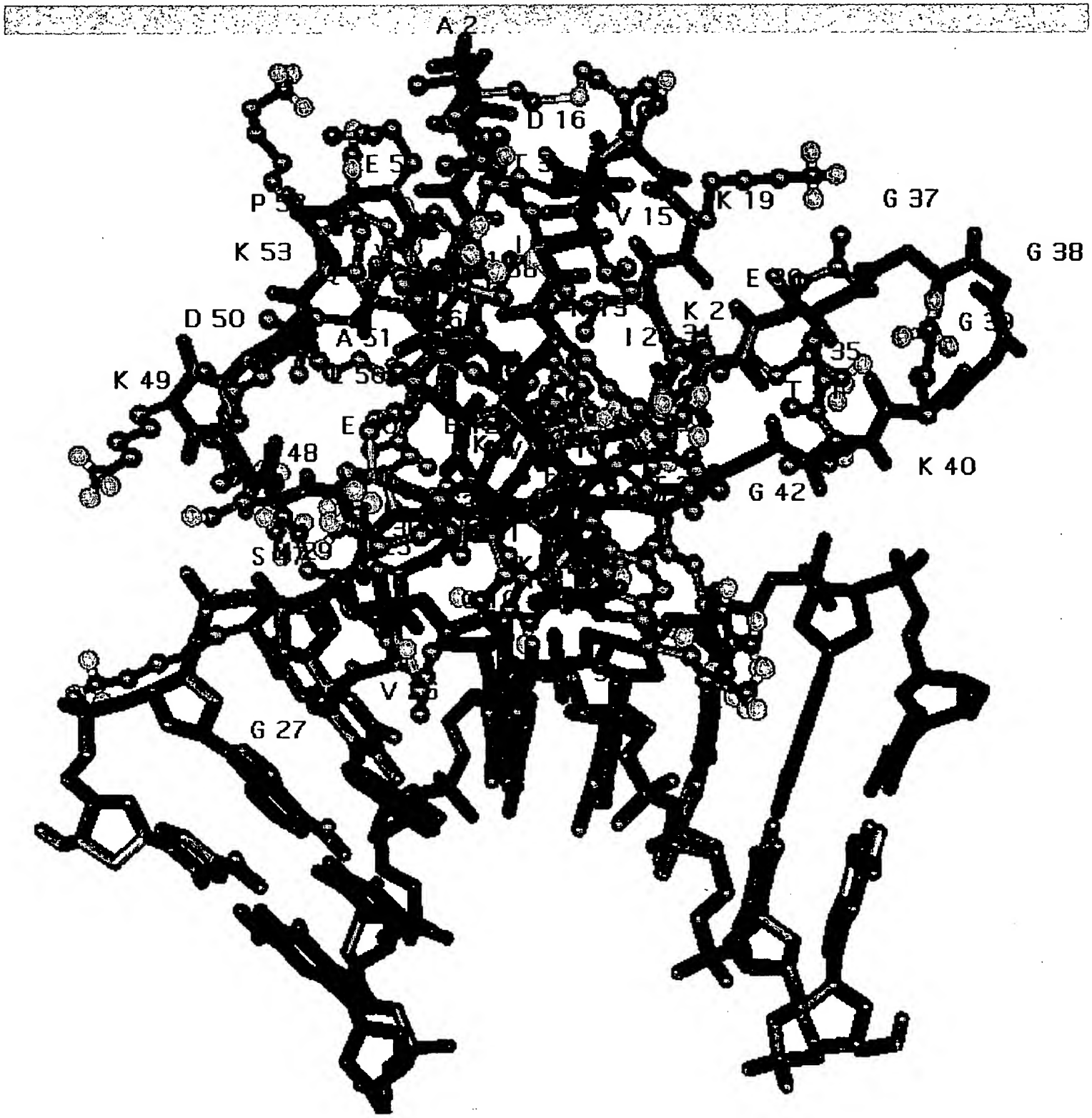


EXHIBIT 5

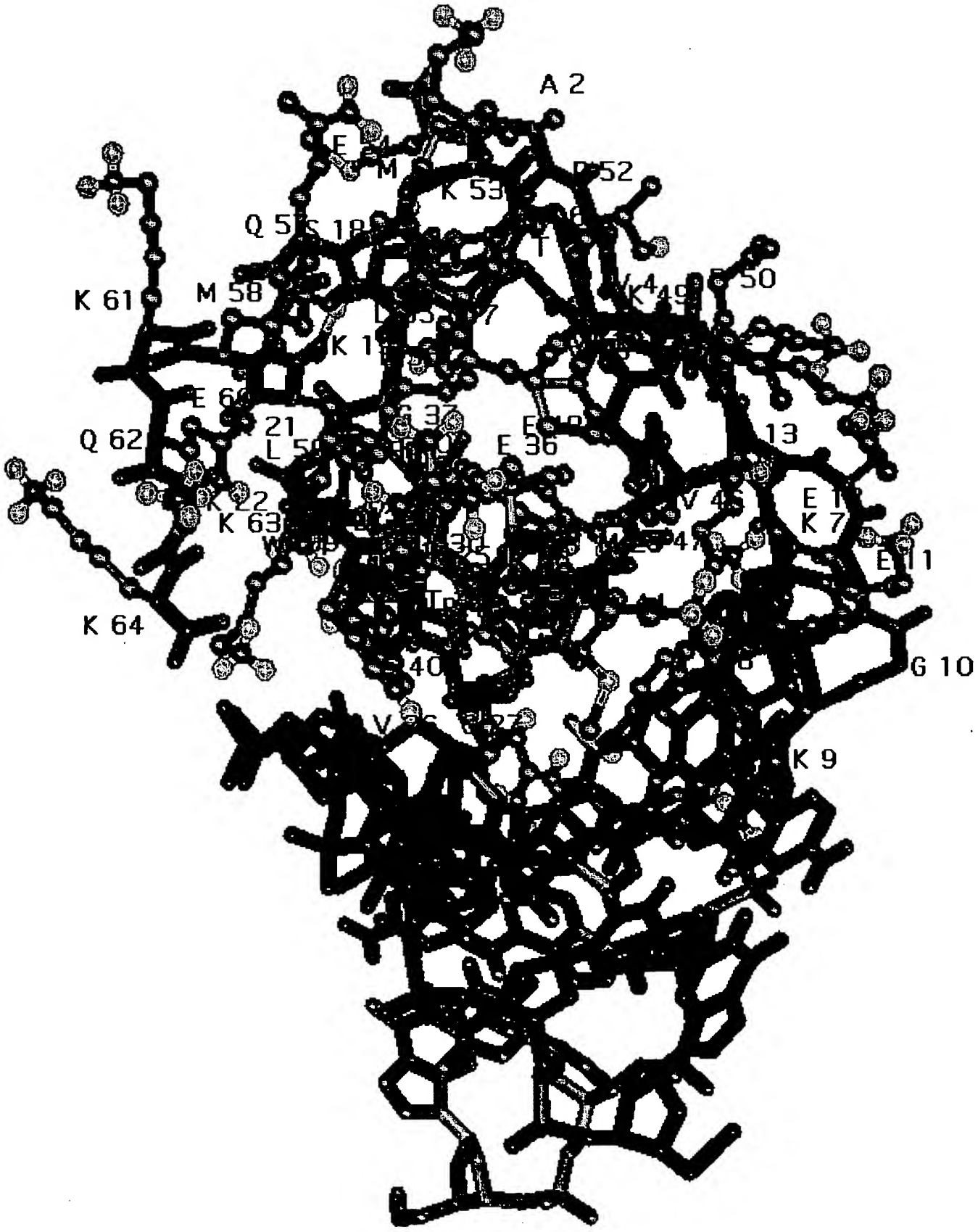


EXHIBIT 6

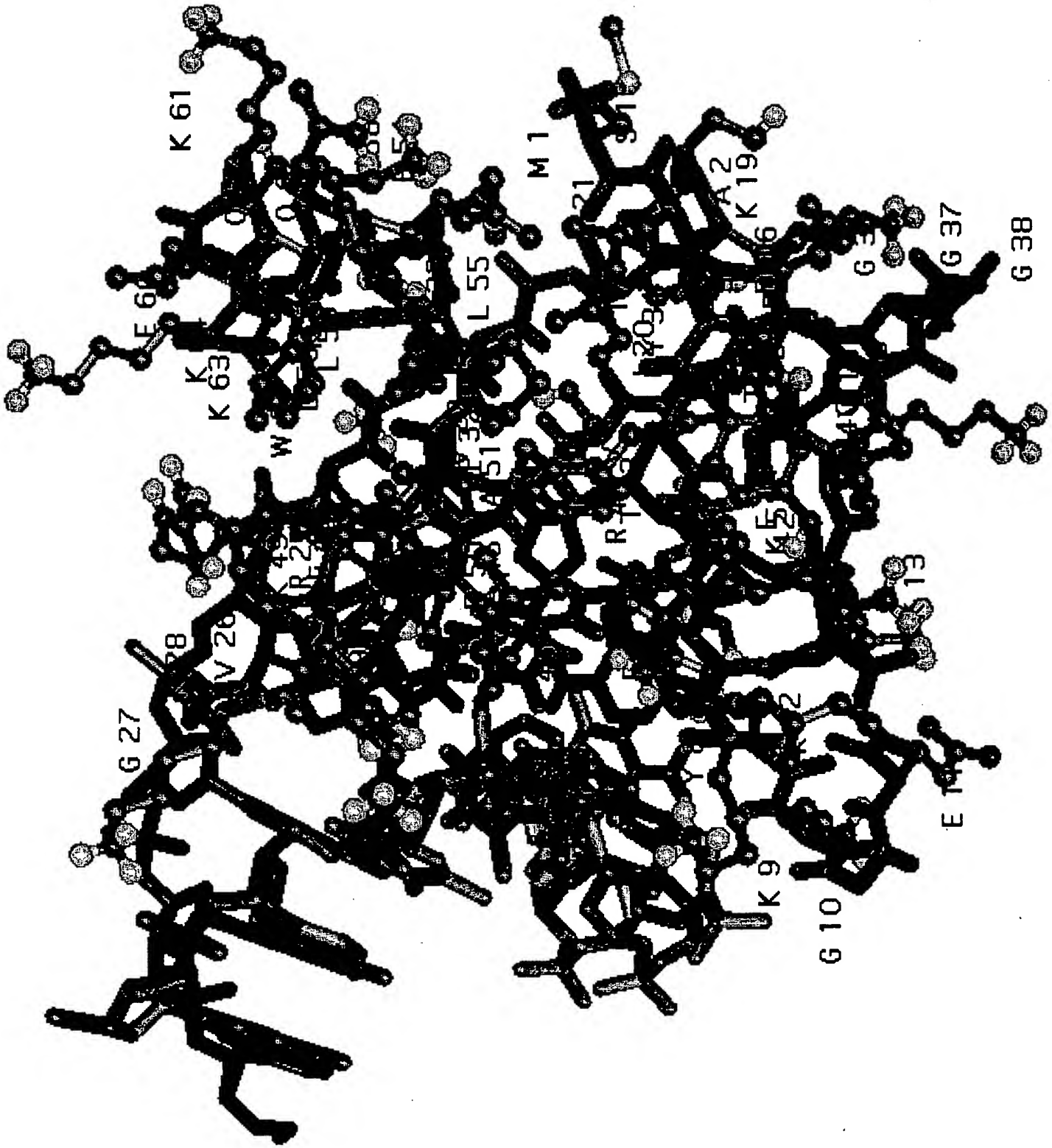


EXHIBIT 7

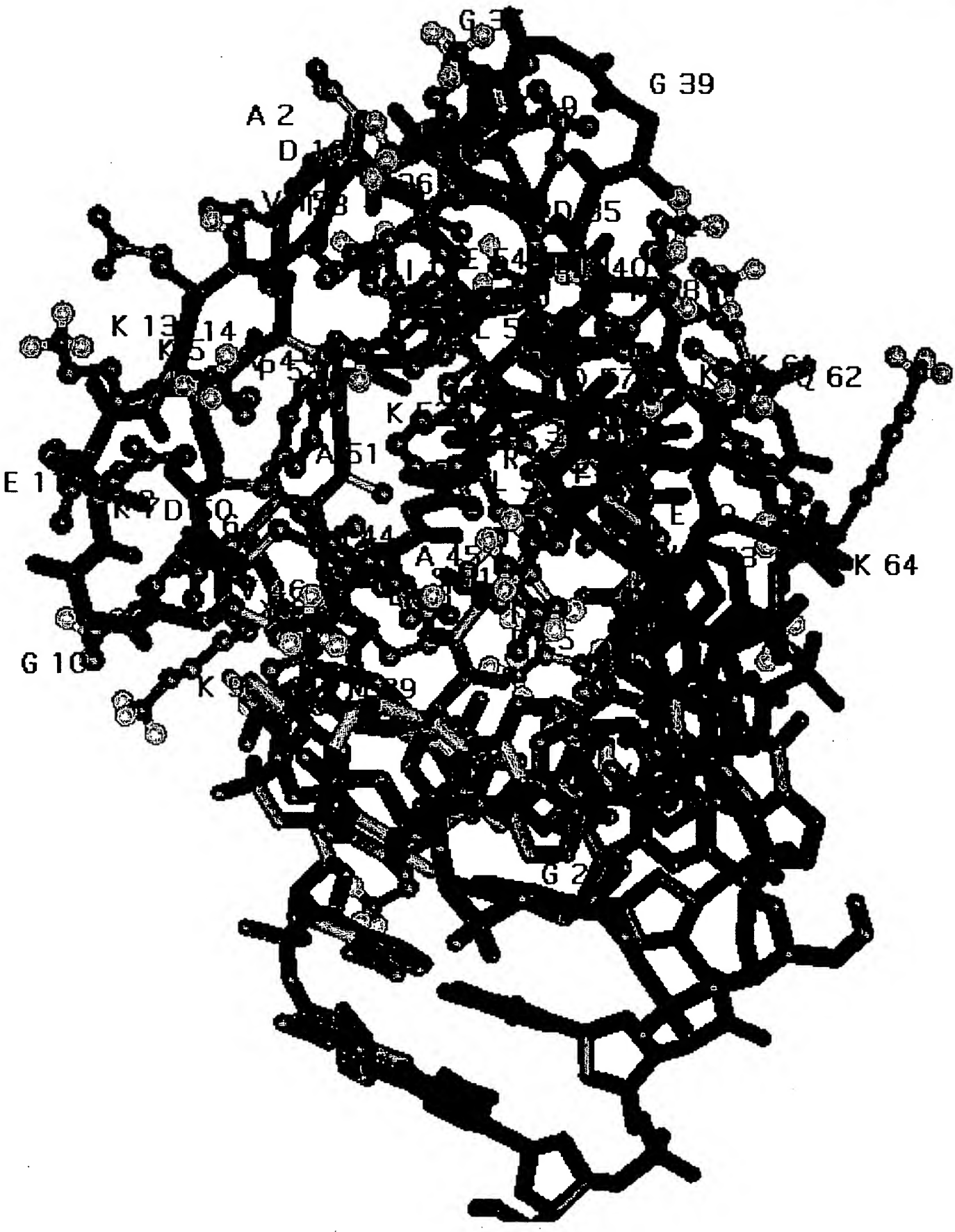


EXHIBIT 8

Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*

Herbert Baumann, Stefan Knapp, Thomas Lundbäck, Rudolf Ladenstein and Torleif Härd

The archaeon *Sulfolobus solfataricus* expresses large amounts of a small basic protein, Sso7d, which was previously identified as a DNA-binding protein possibly involved in compaction of DNA. We have determined the solution structure of Sso7d. The protein consists of a triple-stranded anti-parallel β-sheet onto which an orthogonal double-stranded β-sheet is packed. This topology is very similar to that found in eukaryotic Src homology-3 (SH3) domains. Sso7d binds strongly (K_d < 10 μM) to double-stranded DNA and protects it from thermal denaturation. In addition, we note that ε-mono-methylation of lysine side chains of Sso7d is governed by cell growth temperatures, suggesting that methylation is related to the heat-shock response.

Center for Structural Biochemistry, Karolinska Institutet, NOVUM, S-141 57, Huddinge, Sweden

Correspondence should be addressed to TH

DNA in a random coil conformation occupies a volume that is almost always much larger than the cell in which the molecules are contained. Thus, the DNA of all cells must be structurally organized in a compact form and yet be readily available for transcription. In the nucleus of the eukaryotic cells the genomic DNA is packed by histone proteins into nucleosomes, which in turn form the higher-order structures of chromatin[1]. The structural organization of DNA in prokaryots is somewhat less well understood[2]. Archaea and bacteria contain abundant small, basic proteins which are believed to be involved in packing and unpacking, maintenance and control of the genomic DNA (see refs 2–5 for reviews)—one of the best characterised being the HU protein from *Escherichia coli*. Some of these proteins are also clearly evolutionary related to eukaryotic histones (ref. 6 and work cited therein). Others are believed to have more specialized functions, such as to bend the DNA at specific sites[2].

The thermoacidophilic archaeon *Sulfolobus*, which can be isolated from volcanic hot springs[7], expresses several small, basic proteins. These proteins were first reported by Thomm *et al.* (ref. 8) and were subsequently isolated, characterized and sequenced by Reinhardt and co-workers[9–12]. The basic proteins isolated from *Sulfolobus acidocaldarius* can be grouped into three molecular weight classes of 7,000, 8,000 and 10,000 M_r (7, 8 and 10 kDa), respectively[9,10]. The 7 kDa proteins can be further separated according to their basicity. Sequences are known for the major component of the 7 kDa class in *S. solfataricus* (Sso7d)[12] and the corresponding protein (Sac7d) as well as three minor components (Sac7a, Sac7b, and Sac7e) in *S. acidocaldarius*[9,11]. The sequences of these proteins are compared in Fig. 1*a*. The proteins are all very rich in lysine residues—14 residues out of 63 in Sso7d are lysines. Lysine residues at the amino-and carboxy termini (residues 4, 6, 60, 62 and 63 in Sso7d) are subjected to ε-mono-methylation within the cell[10,12].

The function of the 7 kDa class of proteins in *Sulfolobus* is not known. The initial reports emphasize their DNA-binding properties. The proteins are small, basic and abundant, that is 'histone-like'. Filter-binding assays show that Sso7d binds DNA at physiological salt concentrations and electron micrographs reveal the formation of compacted nucleoprotein particles with both double-stranded (ds) and single-stranded (ss) DNA[12]. The influence of sequence specificity on Sso7d binding to dsDNA has not been investigated. The functional significance of ε-mono-methylation of lysines or the effect of various degrees of methylation on the DNA-binding properties are unclear.

The Sso7/Sac7 class of proteins may also have other functions in addition to DNA binding. For instance, the protein contains the sequence GGGKTGRG (Fig. 1*a*), which is reminiscent of the 'P-loops' found in several classes of ATP- and GTP-binding proteins[13], and might therefore be a phosphate binding site[11].

A protein in *S. solfataricus*, which appears to be identical to the previously identified Sso7d, has been sug-

```
                  10        20        30        40        50        60
Sso7d  ATVKFKYKGEEKQVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDAPKELLQMLEKQ--KK
                * *                                                 *   **
Sac7a  VKVKFKYKGEEKEVDTSKIKKVWRVGKMVSFTYDD-NGKTGRGAVSEKDAPKELLDMLARAEE
Sac7b  VKVKFKYKGEEKEVDTSKIKKVWRVGKMVSFTYDD-NGKTGRGAVSEKDAPKELLDMLA
Sac7d  VKVKFKYKGEEKEVDTSKIKKVWRVGKMVSFTYDD-NGKTGRGAVSEKDAPKELLDMLARAEEK
Sac7e  AKVKFKYKGEEKEVDTSKIKKVWRVGKMVSFTYDD-NGKTGRGAVSEKDAPKELLDMLARAEEK
```

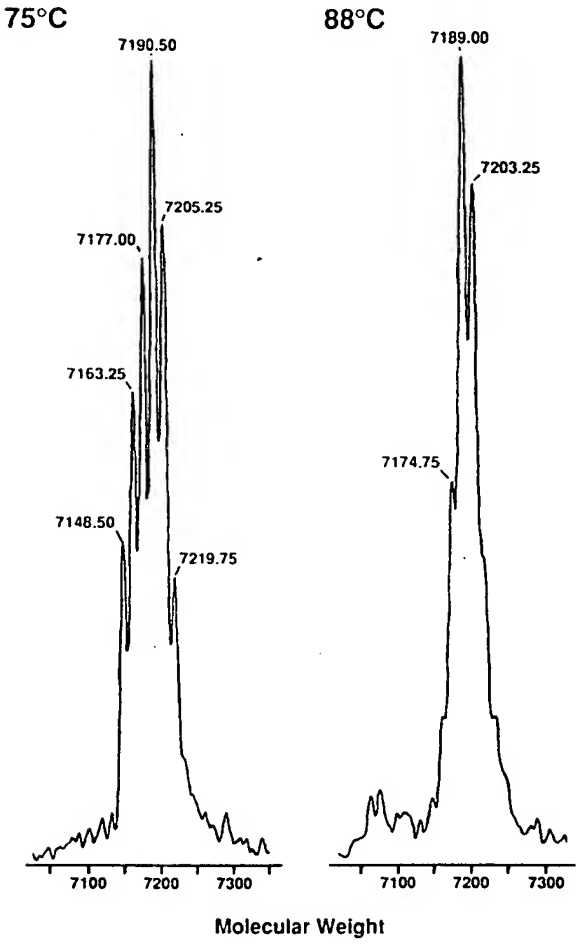


Fig. 1 *a*, Aligned sequences of proteins of basic 7 kDa, proteins from *S. Solfataricus* (Sso7d) and *S. acidocaldarius* (Sac7a,b,d,e)[11,12]. The numbering refers to Sso7d. Stars indicate lysine residues subjected to ε-mono-methylation. The putative phosphate/nucleotide binding site in Sso7d has been boxed. Residue 13 in Sso7d has been changed from Glu[12] to Gln based on NMR data. *b*, Mass spectra of Sso7d from cultures grown at 75°C and 88°C. The numbers indicate calculated masses for the various species. The expected mass for the non-methylated species calculated from the sequence is 7147.2 au.

gested to act as a ribonuclease albeit with a rather narrow substrate specificity[13]. The protein—called p2 by Fusi *et al.*[13], who compare p2 to Sac7d, but seem to have been unaware of the published Sso7d sequence[11]—is reported to be dimeric under native conditions. This observation is in contrast to other data, which clearly show that Sso7d is monomeric (ref. 12, present work).

The abundance of Sso7d in *S. solfataricus*, in combination with its relatively small size, solubility, thermostability, and ease of purification makes the protein suitable for biophysical analyses and structure determination. We have initiated a series of studies to determine the structure and dynamics of the Sso7/Sac7 class of proteins, their nucleic acid–binding affinities and specificities, as well as possible nucleotide binding/hydrolysis. In the present work we report on the structure of Sso7d in solution and provide a more detailed characterization of its DNA-binding properties. When analyzing the structure of Sso7d we made the intriguing observation that this abundant archaeal protein in fact is structurally similar to that of SH3 domains involved in signal transduction in eukaryote. We also note that the extent ε-mono-methylation of lysine residues in Sso7d depends on cell culture growth temperature, suggesting that the methylation is a response to heat shock.

Purification and initial characterization

Sso7d was purified from *S. solfataricus* (Methods); the protein eluted in two peaks from the Mono S column used in the final purification step. Mass spectrometric analysis of (pooled) material from the two peaks indicate the presence of six masses (Fig. 1*b*). Mass differences correspond to sequential substitutions of hydrogen atoms with methyl groups, as a result of the ε-mono-methylation of lysine residues described previously[11,12]. The observation of six peaks with different methylation patterns is consistent with the notion that five lysine residues are subjected to methylation. The mass of the species with the lowest molecular weight corresponds with that calculated from the sequence (Fig. 1*a*).

Sso7d from the two fractions show NMR chemical shift differences of 0.02–0.12 p.p.m. affecting backbone resonances of residues 2, 3, 6, 11, 12, 16, 17 and 44, but connectivities involving these residues observed in 2D NOESY spectra are practically identical for material from the two fractions. The chemical shift differences are most likely caused by electrostatic effects due to methylation of one of the lysine residues, because differences in chemical composition can be ruled out based on mass spectrometry. The presence of two exchanging conformers can also be ruled out because NOESY spectra recorded on the two (separated) species (samples 2 and 3

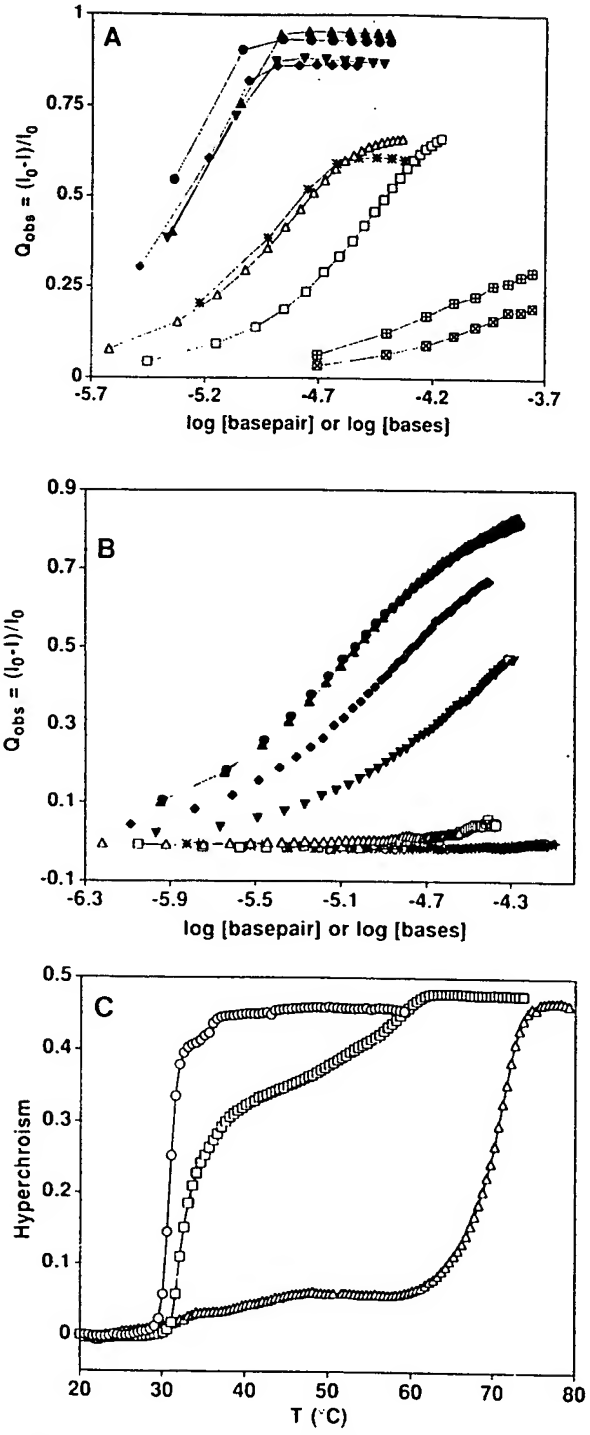


Fig. 2 Analysis of DNA binding by Sso7d. *a*, Equilibrium titrations of Sso7d with various polynucleotides and monodinucleosides based on fractional fluorescence quenching (Q_{obs}). The titrations are performed at low salt concentration (buffer D) as reverse titrations in which the protein concentration is kept constant (2μM). *b*, Equilibrium titrations performed at a higher salt concentration, which is closer to physiological conditions (buffer C) with 1 μM protein. Symbols in *a* and *b* refer to titrations with poly(dGdC) (1), poly(dAdT) (▼), poly(dIdC) (●), poly(dAdU) (▲), poly(dA) (△) poly(dC) (□), poly(rA) (,), poly(rC) (+), dATP (⊞)and dCTP (⊠). The abscissa legends indicate that concentrations of double-stranded DNAs are measured in base pairs and concentrations of single-stranded polynucleotides and monodinucleosides are measured in bases. *c*, Thermal denaturation profiles of poly(dIdC) in the absence and presence of bound Sso7d; no added protein (○), Sso7d added to a concentration corresponding to 1:15 Sso7d:DNA bp (□), and Sso7d added to a concentration corresponding to 1:3.6 Sso7d:DNA bp (△). The poly(dIdC) concentration was 12 μM bp. The thermal denaturation experiments were performed at low salt concentration conditions (buffer E).

see Methods) do not change within a period of several months.

The extent of ε-mono-methylation of lysine side chains varies with bacterial growth conditions so that higher growth temperatures lead to more extensive methylation (Fig.1*b*). The physiological relevance of this effect is not clear. It is possible that the lysine methylation is directly related to the stability of the protein and/or the DNA-protein complex and the response of the organism to heat shock. The pK_a of the lysine side chain is affected very little by methylation[16] and it seems less likely that methylation has a direct affect on DNA-binding affinity.

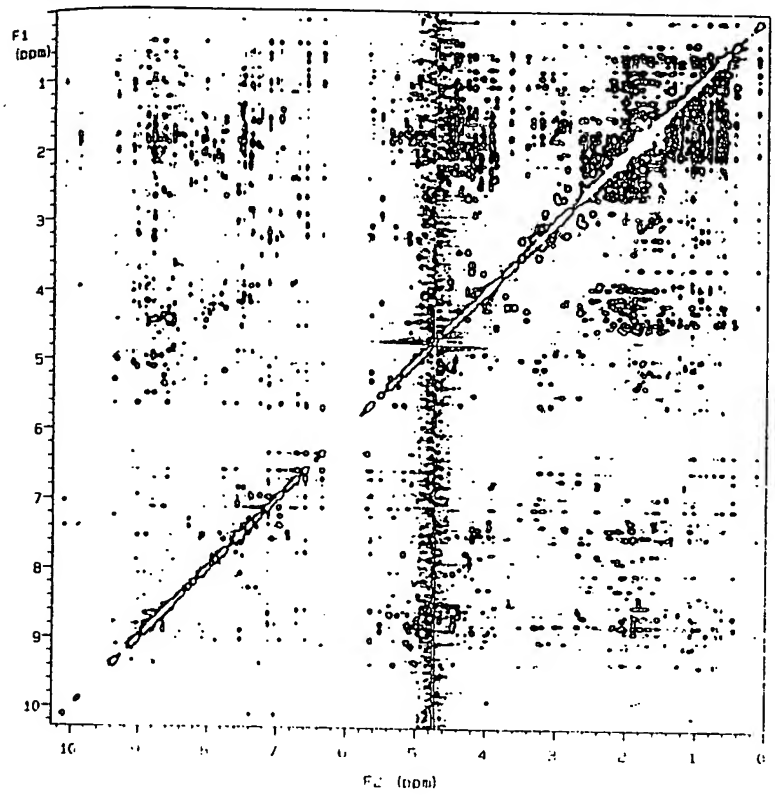
Sso7d binds strongly to dsDNA

The equilibrium binding of Sso7d to various polynucleotides was studied by monitoring changes in the intrinsic tryptophan fluorescence on formation of the complexes. The fluorescence of Trp 23, excited at 290 nm, is quenched by 60–90% on binding and the emission spectrum is shifted to longer wavelengths (not shown).The results of titrations performed at low salt (buffer D) and physiological salt concentration (buffer C) conditions, respectively, are shown in Fig. 2*a,b*. Titration curves for four different dsDNA polymers with alternating purine-pyrimidine sequences at low salt, show an observed quenching, Q_{obs}, which levels out at Q_{obs}~0.9. There is little difference in the apparent binding affinity to the various dsDNAs at low salt, presumably due to quantitative binding to all DNAs. The binding curves show saturation at an approximate concentration ratio of 1:6 protein:DNA base pairs (bp), which can be taken as an estimate of the lower limit for the Sso7d binding site density on DNA.

There is a definite difference between the Sso7d binding affinities to various dsDNA sequences at physiological salt concentrations (Fig. 2*b*). The binding is strongest to poly(dIdC) and poly(dAdU), for which the affinities are approximately equal. The DNA concentration at half saturation is in this case approximately 8 μM bp. This number corresponds to an affinity constant of ~$0.5\text{-}1 \times 10^6$ $(M \text{ sites on DNA})^{-1}$ if one (conservatively) assumes that the maximum binding site density is in the range 1:6–1:3 protein:DNA bp. Binding to poly(dGdC) is somewhat weaker and binding to poly(dAdT) is about 5–10 times weaker than that to poly(dAdU) and poly(dIdC).

The binding affinities of Sso7d to various alternating dsDNA sequences can be rationalized as follows. First, a

a



b

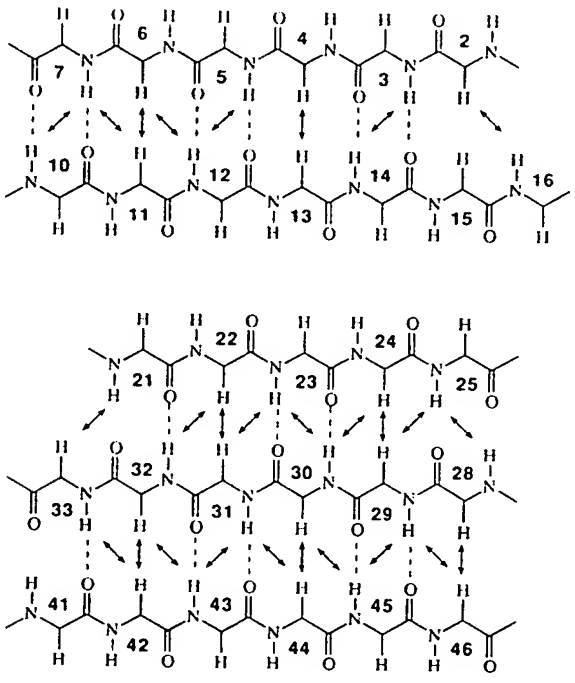


Fig. 3 *a*, Two-dimensional 500 MHz NOESY spectrum of Sso7d (concentration ~2.5 mM in 90%:10% H_2O:D_2O). *b*, Schematic view of the two antiparallel β-sheets in Sso7d. Hydrogen bonds used in the SA simulations and observed NOEs are indicated with dashed lines and arrows, respectively. Additional hydrogen bonds, not used in SA, are described in the text.

methyl group at position 5 (in the major groove) of the pyrimidine is unfavourable for binding. This is clear when comparing binding to poly(dAdU) and poly(dAdT). Thus, DNA-protein interactions may occur within the DNA major groove. Second, binding to dsDNA sequences with two inter-strand hydrogen bonds is stronger than to those with three hydrogen bonds in polymers lacking the pyrimidine methyl (that is, when comparing poly(dAdU) and poly(dIdC) to poly(dGdC)). This behaviour might be related to some physical property such as flexibility, considering that Sso7d seems to induce condensation of DNA[4].

Titration curves for Sso7d binding to ssDNA and ssRNA homopolymers in the presence of low salt concentrations show saturation at Q_{obs}=0.6–0.7. The binding to ssDNA and ssRNA under these conditions appear to be weaker than that to dsDNA, although there is a possibility that these complexes are as strong as those with dsDNA but that the maximum binding-site density is lower. However, the thermal denaturation studies described below indicate that dsDNA is preferred over ssDNA, because the melting temperature increases on formation of the complex. Furthermore, increasing the salt concentrations to physiological levels has a dramatic effect on the binding to single-stranded polynucleotides (Fig. 2*b*). Under these conditions there is only very weak binding to poly(dA) and poly(dC), whereas no binding to poly(rA) and poly(rC) can be detected at polymer concentrations <100 μM bases. Thus, there seems to be a large binding preference for dsDNA compared to ssDNA and ssRNA at higher salt concentration conditions.

At low salt concentrations it is also possible to monitor binding of the monodeoxynucleosides dATP and dCTP through the quenching of Trp 23 fluorescence (Fig. 2*a*). The titration curves do not show saturation and it is difficult to estimate stoichiometries and affinities based on the present data, but the binding seems to be weaker than that of the DNA and RNA polymers.

Protection of DNA from denaturation

Thermal denaturation profiles of double-stranded poly(dIdC) in the absence and presence of bound Sso7d are shown in Fig. 2*c*. Poly(dIdC) is thermally unstable above 32°C at the conditions used in the experiment shown in Fig. 2*c*. Addition of less than stoichiometric amounts of Sso7d increases the thermal stability of poly(dIdC) yielding a biphasic DNA melting curve. Saturation of poly(dIdC) with bound Sso7d again results in a single phase denaturation profile with a melting temperature of about 70°C. Thus, binding of Sso7d increases the melting temperature of poly(dIdC) by more than 38°C at low salt concentrations. Similar, albeit somewhat attenuated, effects can be observed with shorter DNA oligomers at physiological salt concentrations (data not shown). It is difficult to quantify the effect of Sso7d binding to DNA polymers at high salt concentrations because melting temperatures are high even in the absence of bound protein. However, it seems possible that Sso7d binding may shift the melting temperature of DNA above that of the boiling point of water.

The remarkable effect of Sso7d binding on DNA thermal stability is very similar to that of the HTa protein from *Thermoplasma acidophilum*[17]. Stein and Searcy[17] argue that the HTa protein may act to protect bacterial DNA during short periods of denaturing conditions allowing the organism to cope with transient periods of high temperatures.The Sso7d protein may function in a similar manner in *Sulfolobus*. The different extent of lysine methylation of proteins expressed at different growth temperatures may also relate to the bacterial response to heat shock and stabilization of functionally important proteins. However, the effect of Sso7d methylation on its DNA-stabilizing properties are unknown.

NMR structure determination

Two-dimensional NMR spectra of Sso7d were recorded at 500 and 600 MHz. The ^{1}H spectrum (Fig 3*a*) shows a very favourable resonance dispersion and could be almost completely assigned using standard methodologies[18,19]. Upon assigning the sequence we found one disagreement with the published sequence: residue 13, which is a Glu in the sequence of Choli *et al.*[12], is in fact a Gln and this correction has been made in Fig. 1*a*. The ^{1}H linewidths in Sso7d (3–8 Hz) are typical for a protein with a relative molecular mass of 7,000, indicating that Sso7d is predominantly monomeric under the conditions used in the NMR experiments

The NOESY spectrum of Sso7d contains stretches of very strong sequential $d_{\alpha N(i),N(i+1)}$ NOE connectivities in combination with strong long range $d_{\alpha\alpha(i,j)}$ and $d_{\alpha N(i),N(j+1)}$ NOEs, which are typical for β–sheet secondary structures[18]. These arise from one double-stranded and one triple-stranded anti-parallel β–sheet (Fig. 3*b*). The pattern of intra- and inter-residue NOE connectivities, the observation of slowly exchanging backbone amide protons and low amide temperature coefficients allowed the identification of 14 intramolecular backbone-backbone amide hydrogen bonds within the anti-parallel β–sheets (Fig. 3*b*).

The three-dimensional structure of a fragment containing residues 1–62 of Sso7d was calculated using a dynamic simulated annealing (SA) protocol with 617 non-redundant NOE distance constraints, 11 χ^1 dihedral–angle constraints and 28 hydrogen bond distance constraints (two constraints per hydrogen bond), that is 10.6 constraints per residue. The NOE distances (d_{ij}) were distributed as 233 intraresidue (i=j), 151 sequential ($|i-j|=1$), 51 medium range ($2\leq|i-j|\leq4$), and 182 long range ($|i-j|\geq5$) NOEs (Table 1). The quality of the computed SA structures is good as judged from the low Lennard-Jones potential energies and the very small average deviations from idealized geometries. The distance constraint violation statistics are also good: the average number of distance constraint violation >0.3 Å is 0.2 per structure and the largest violation found in any of the 35 structures is 0.38 Å. The largest dihedral angle constraint violation is 3.2°.

A plot of average backbone dihedral angles in the 35 SA structures is shown in Fig. 4*a* and plots of dihedral angle order parameters are shown in Fig. 4*b–d*. Average backbone dihedrals are all within the allowed regions of a Ramachandran diagram (not shown), except for those of Lys 8. The backbone of this residue is less well defined, as judged from the angular order parameters, which results in a sterically unfavourable geometric average. The superimposed backbones of the final SA structures are shown in stereo in Fig. 5*a*. The backbone conformation within the β–sheet regions is well-defined, as indicated by atomic backbone root–mean–square devia-

Table 1 Structural Statistics for Sso7d[a]

Statistic	<SA>	$(SA)_{av\,rmin}$
R.m.s. deviation from experimental distance (Å) and dihedral angle (deg) restraints[b]		
distance restraints (617)	0.025 ± 0.0018	0.024
dihedral angle restraints (11)	0.26 ± 0.23	0
No. of violations[c]		
distance restraints (>0.3 Å)	0.20	0
dihedral angle restraints (>1°)	0.31	0
$E_{L,J}$ (kcal mol^{-1})[d]	-172 ± 20	-214
Deviations from idealized covalent geometry		
bonds (Å)	0.0025 ± 0.00016	0.0026
angles (deg)	0.36 ± 0.015	0.36
impropers (deg)	0.24 ± 0.03	0.22

[a] The notation of the NMR structures is as follows: <SA> are the final 35 simulated annealing structures; $(SA)_{av\text{-}rmin}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best fit to each other followed by minimization by restrained regularization.
[b]The number of restraints is given in parentheses.
[c] The maximum distance violation is 0.38 Å and the maximum dihedral angle violation is 3.2° in an individual SA structure.
[d] $E_{L,J}$ is the Lennard-Jones van der Waals' energy calculated with the CHARMM[47] force field.

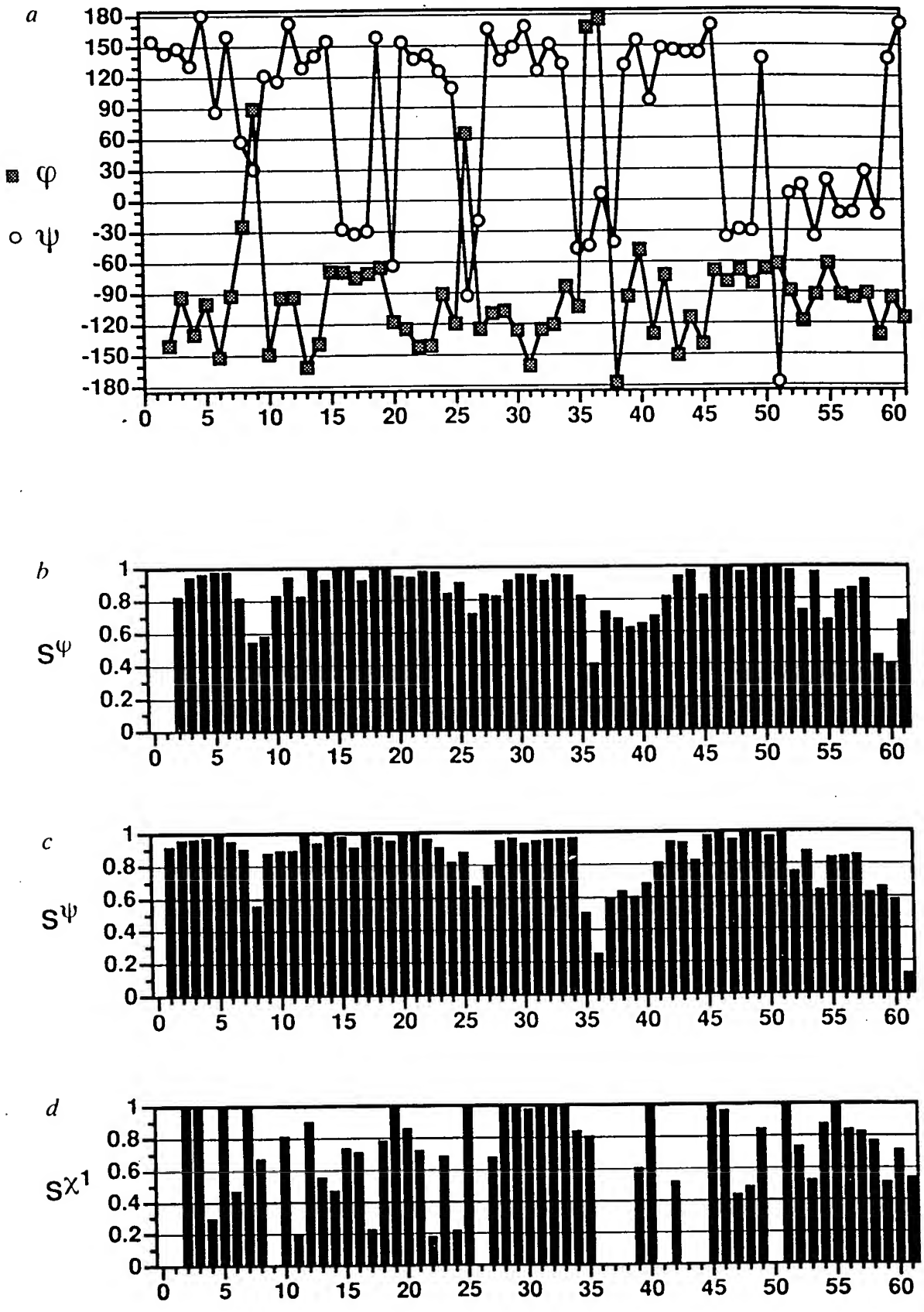


Fig. 4 Average φ and ψ dihedral angles (*a*) and angular order parameters S^{angle} for φ (*b*), ψ (*c*) and χ^1 (*d*) dihedral angles for all residues in the 35 SA structures.

tions of 0.5±0.1 Å compared to the geometric average structure (Table 2). Other regions are somewhat less well defined, as indicated by an overall backbone r.m.s.d. of 1.1±0.2Å. The side chains of several residues in the hydrophobic core of Sso7d are also well resolved, as can been seen in Fig. 5*b*. The C-terminal fragment (residues 46–60) is somewhat more well defined than the loop regions, with a backbone r.m.s.d. of 0.9±0.2 Å, and a short α-helix including residues 52–59 is clearly discernible. This helix can also be deduced from a continuous stretch of strong sequential $d_{NN}(i,i+1)$ and medium range $d_{\alpha N}(i,i+3)$ and $d_{\alpha\beta}(i,i+3)$ NOE connectivities.

The final set of SA structures contains several hydrogen bonds, in addition to those used in the structure calculations. These involve the backbone amide protons and carbonyl oxygens of residues 18 and 15, 19 and 15, 20 and 32, 25 and 28, 27 and 25, 50 and 46, and 50 and 47, respectively.

The Sso7d structure

Sso7d is a globular protein. The tertiary fold consists of a triple-stranded anti-parallel β–sheet, consisting of residues 21–25, 28–33 and 41–46 (strands III, IV and V, respectively), onto which a double-stranded β–sheet, made up of residues 2–7 and 10–15 (strands I and II), is packed in an orthogonal manner. The hydrophobic core consists of side chains at the interface of the two sheets, including those indicated in Fig. 5*b*. Strands I and II are connected through a type II reverse turn with a hydrogen bond between the carbonyl of Tyr 7 and the amide of Glu 10. Strand II ends in one complete turn of an α–helix involving residues 16–19, with a hydrogen bond between the carbonyl of Asp 15 and the amide of Ile 19. Strands III and IV in the second β–sheet are connected by a type I reverse turn involving residues 25–28. Thus, hydrogen bonds between the carbonyl of Val 25 and the amide of Met 28, and the amide of Val 25 and the carbonyl of Met 28 are present in the triple-stranded β–sheet, in addition to those shown in Fig. 3*b*. Residues 35–40 form a surface loop, containing the glycine tripeptide Gly 36-Gly 37-Gly 38 (Fig. 6). The structure of this loop is not very well defined by the NMR constraints and it is clear that it can show a large degree of inherent flexibility. Strand V (residues 41–46), ends in a complete turn of an α–helix involving residues 47–50. This short helical segment is anchored through hydrophobic interactions involving Ala 50 and Pro 51. The backbone of the C-terminal fragment is not as well-defined as the β–sheets, but residues 52–59 appear to form two turns of α–helix. This short helix is packed against the core through hydrophobic interactions between Leu 54 and Ala 50.

The surface of Sso7d contains a hydrophobic cleft and several exposed hydrophobic side chains (Fig. 6*a*). The hydrophobic cleft consists of the N-terminal Ala 1 side chain and the isoleucine residues Ile 16 and Ile 19 on one 'side', and the side chains of Pro 51, Leu 54 and Met 57 of the C-terminal helix on the other. The Trp 23 and Val 25 side chains of strand III are completely exposed to the solvent and so is the methyl of Ala 44. The side chains of Tyr 7 and Met 28 are partially exposed on the surface.

The many basic lysine and arginine side chains are rather evenly distributed at the surface and the positive charges seem to be partially compensated for by nearby acidic side chains. However, the face of the triple-stranded β–sheet appears to be predominantly positive in charge. This surface also contains the exposed Trp 23 side chain: the fluorescence of this residue is quenched by 90% upon formation of a complex with dsDNA. Thus, this face of the protein may be the DNA binding surface.

Sso7d and eukaryotic SH3 domains

The topology of Sso7d is very similar to that of eukaryotic SH3 domains (Fig. 7*a*). The SH3 domains are small protein modules(about 60 residues) which, together with SH2 domains, are found in many proteins involved in signal transduction in eukaryote[20]. The SH2 and SH3 domains are commonly found in kinases or phospholipases, where they are believed to participate in protein-protein interactions. The structures of SH3 domains from several proteins have recently been solved by both NMR spectroscopy and X-ray crystallography[21,22].

The minimized average structure of Sso7d is compared with the structures of the SH3 domains of chicken brain α spectrin[23] (PDB entry 1SHG) and human *fyn* proto-oncogene[24] (PDB entry 1SHF) in Fig. 7*a* and an alignment of the three sequences based on secondary structure and folding topology is shown in Fig. 7*b*. The superimpositions included 38 Cα coordinates of the five β-strands and a fragment from the C terminus in Sso7d (residues 1–7, 10–16, 21–25, 28–33, and 41–53; Fig. 7*b*). The r.m.s.ds with corresponding fragments in the α spectrin and fyn SH3 domains are in both cases 3.3 Å. Thus, there is a good quantitative agreement between these structures. Differences are found at the N and C termini and for surface loops. In particular, the interconnection between the β–strands of the two SH3 domains which corresponds to strands IV and V in Sso7d is extended into the putative P-loop in Sso7d (Fig. 7*a*).

Comparison of the complete sequences of Sso7d and the SH3 domains does not reveal sequence homology. However, homology can be inferred when considering only the fragments for which there is structural similar-

Table 2 Atomic r.m.s. difference statistics for the Sso7d structure[a]

Comparison	Residues	Backbone[b]	All heavy atoms
<SA> vs SA_{av}	1-60	1.08±0.17	1.60±0.16
	46-60	0.95±0.22	1.72±0.28
	2-7,10-15,21-25, 28-34,41-45	0.54±0.09	1.14±0.11
SA_{av} vs $SA_{av\text{-}rmin}$	1-60	0.45	0.80

[a]Notations correspond to those defined in Table 1 with the addition that SA_{av} is the non-minimized geometric average structure. Residues 61 and 62 are excluded from the comparison due to lack of structural constraints in this region.
[b]Superimposed fragments.
[c]Atoms N, C and Cα.

ity, that is, when excluding loops and N and C termini, although any homology is still too weak to be conclusive by conventional alignment algorithms. Sequence identities and sequence similarities (aromatic/hydrophobic residues) in the fragments that were used in the structural alignment are shown in Fig. 7*b*. It is worth noting that several residues which are well conserved among various SH3 domains[22] are present at the corresponding positions in Sso7d. These include Val 3 in Sso7d (an alanine in SH3), Phe 5 and Tyr 7 (aromatics), Lys 12 (lysine), Val 22 and Trp 23 (hydrophobic), Met 28 and Ile 29 (tryptophan and tryptophan/hydrophobic), Gly 43 (glycine), Ala 44 and Val 45 (aromatic or hydrophobic), and Ala 50 (hydrophobic). Sso7d and SH3 domains are also similar in that they expose hydrophobic surfaces[21].

The possible origin and significance of the structural similarity between the Sso7d, which is an abundant protein in the archaeon *Sulfolobus*, and the SH3 domains, which appear to have assumed highly specialized roles in signal transduction in eukaryote, is unclear. One scenario may be that the fold has survived in all kingdoms due to its (thermal) stability and because it forms a suitably small and stable platform for different functions in various organisms. An SH3-like fold has also recently been discovered for a small protein in the photosystem I complex (PsaE) in cyanobacteria[23]. Structural similarities to SH3 has also been noted in another DNA-binding protein: the biotin biosynthetic operon repressor (BirA) in *E. coli*[24].

Methods

Protein purification. *Sulfolobus solfataricus* (DSM 1617) isolated from volcanic hot springs in Italy[17] was purchased from the

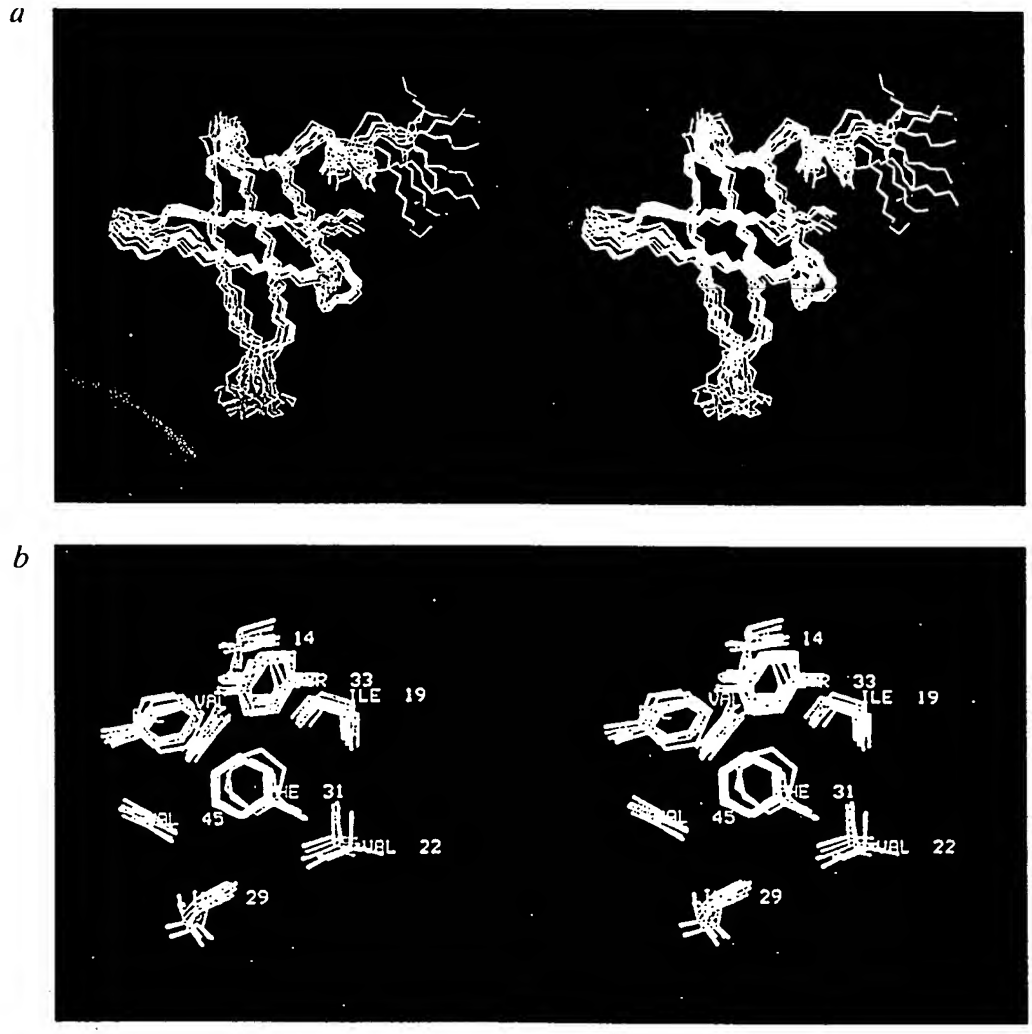


Fig. 5 *a*. Stereoview of superimposed backbone traces of residues 1–62 in Sso7d. For the sake of clarity, only 11 of the 35 SA structures are shown. The structures are superimposed to minimize r.m.s differences of backbone atoms in residues 1–60. N and C termini are coloured in blue and red, respectively. The loop containing the putative phosphate/nucleotide binding site is coloured in green. *b*. Stereoview showing the resolution and packing of hydrophobic side chains in the protein core. The structures have in this case been superimposed to minimize r.m.s. deviations between heavy atoms of residues constituting the core.

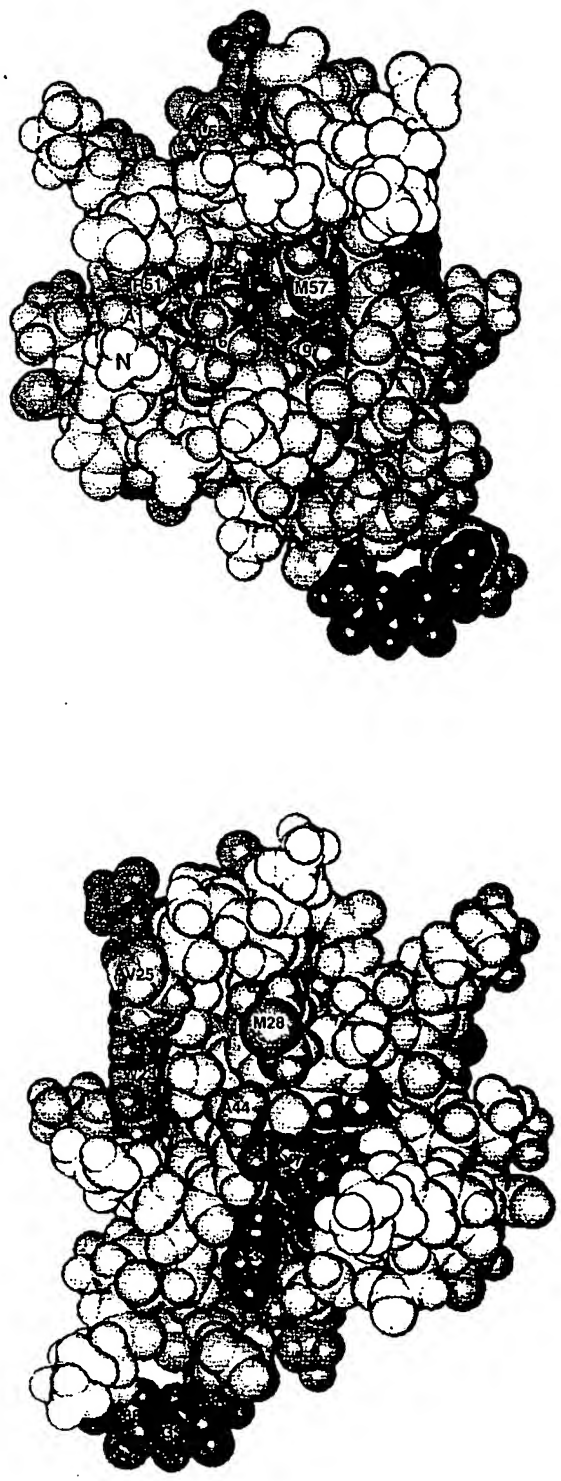


Fig. 6 Space-filling model of Sso7d showing exposed hydrophobic (yellow) and aromatic (orange) side chains (tyrosine hydroxyls are also coloured in orange). The glycines in fragment 36–38 are coloured in green. The views in (*a*) and (*b*) are from opposite directions. N and C termini are indicated in (*a*).

"Deutsche Sammlung von Mikroorganismen" (DSM, Braunschweig). Cultivation was performed aerobically at 75°C (re: 7) with an additional 10 gl^{-1} saccharose in a membrane fermente (Bioengineering). The cells were heat-shocked for 90 min at 88°C and harvested by centrifugation. Protein was also purified from cells that had not been subjected to heat shock, for comparison of the extent of lysine methylation.

100 g cells were lysed in buffer A (10 mM Tris buffer, pH 8.8, wit 20 mM NaCl, 10% Glycerol) by passing the cell suspension throug a French press. The lysate was centrifuged to remove cell debris and dialyzed against the same buffer. The cytosolic proteins were loaded onto a Mono Q (Pharmacia HR10/10) column equilibrated with buffer A: Sso7 was found in the flow-through. This fraction was concentrated in an Amicon stirred cell and applied in 1.5 m fractions to a Superose 6 column (90 x 1.5 cm) equilibrated wit 30 mM Tris/HCl and 200 mM NaCl at pH 7.4 . Fractions containing Sso7 were pooled, dialyzed against 50 mM potassium phosphate 50 mM NaCl at pH 6.0, loaded onto a Mono S (Pharmacia HR10 10) column equilibrated with the same buffer and eluted with a linear gradient of buffer B (50 mM potassium phosphate pH 8, 1 NaCl). Sso7d eluted at 25% B in two separate peaks, due to the presence of differently methylated species of the protein.

Sso7d concentrations were measured spectrophotometrically on a Cary 4E spectrophotometer using an extinction coefficient calculated from tyrosine (ε_{280nm} = 1400 M^{-1} cm^{-1}) and tryptophan (ε_{280nm} = 5500 M^{-1} cm^{-1}) absorption[24].

NMR samples were prepared in 90%: 10% H_2O: D_2O or 100% D_2O with 20 mM potassium phosphate (pH 5 or 6), 50 mM NaCl and 0.1% azide. The structure determination is based on data recorded on the following four NMR samples: 2.5 mM protein at pH 6 containing material from both peaks eluted from the Mono S column.; ~0.2 mM protein at pH 6 containing material eluting under peak 2; 1 mM protein at pH 5 containing material eluting under peak 1; and 2 mM protein containing both fractions in D_2O buffer at pH 6 (non-corrected pH meter reading). The first and last samples contained two distinct NMR species. A combination of spectra collected on the second and third samples corresponds to the NMR spectrum of sample 1.

Mass spectrometric analysis. Mass spectrometry (MS) was carried out at Pharmacia Bioscience Center, Stockholm, using a VG Platform mass spectrometer from Fisons Instruments equipped with an electrospray interface. The mobile phase consisted of methanol:water (1:1) with 1% acetic acid. The range 700 < (M/z) < 1700, where M is the mass and z is the charge was scanned and calibrated using horse heart myoglobin as a calibration standard. Uncertainties in molecular mass determinations are approximately two mass units.

Equilibrium titrations. The DNA and RNA polynucleotides used were purchased from Pharmacia and dissolved in 150 mM NaCl and 10 mM Tris/HCl at pH 7.4. Polynucleotide concentrations were determined spectrophotometrically using extinction coefficients given by Pharmacia. The deoxynucleosides ATP and CTP were purchased from Boehringer-Mannheim.

Equilibrium titrations were carried out at 20°C in buffer C (100 mM NaCl, 1 mM $MgCl_2$, 0.1 mM octaethylene glycol monododecyl ether ($C_{12}E_8$) and 20 mM Tris/HCl at pH 7.4) and in buffer D (0.5 mM $C_{12}E_8$ and 20 mM Tris/HCl at pH 7.4), for which the pH measurements refer to 20°C. Titrations were performed as reverse titrations, in which different amounts of DNA/RNA were added at constant protein concentration (1 μM in buffer C and 2μM in buffer D). Steady-state fluorescence measurements were carried out on a Shimadzu RF-5000 spectrofluorophotometer using the methodology and additional titration instrumentation recently described elsewhere[25]. The excitation wavelength was 290 nm and emission intensities were sampled at 0.2 nm intervals within the wavelength range 340–355 nm. Emission spectra were recorded five times for each titration point in order to minimize effects of instrumental fluctuations. Measured fluorescence intensities were corrected for background emission by subtracting (small) signals from buffer samples and for optical filtering effects due to DNA absorption at 290 nm.

The fractional fluorescence quenching (Q_{obs}) was calculated as $(I_0-I)/I_{\infty}$, where I_0 is the protein fluorescence intensity observed in the absence of DNA/RNA and I is the intensity in the presence of DNA/RNA. Binding isotherms are presented as plots of Q_{obs} against the logarithm of the basepair (dsDNA) or base (ssDNA, ssRNA and monodeoxynucleosides) concentration.

DNA melting studies. Light absorption of poly(dIdC) at 260 nm was measured as a function of temperature on a CARY 4E spectrophotometer, which allows the simultaneous measurement of up to six melting curves. The temperature was increased in steps of 1°C during a time period of 30 s, followed by a holding time of 60 s prior to absorbance measurements. The denaturation experiments were performed in 5 mM Tris/HCl at pH 7.0 (buffer E) with various concentrations of added Sso7d.

NMR spectroscopy. NMR spectra were recorded on Varian Unity 500 and 600 NMR spectrometers operating at magnetic fields of 11.74 and 14.09 T, respectively, and equipped with programmable pulse modulators and pulsed field gradient hardware. Spectra were recorded at 293, 303, 313 and 323 K. ^{1}H chemical shifts at 303 K (available from the authors) are referenced to H_2O at 4.74 p.p.m.. Phase sensitive two-dimensional spectra were recorded in the hypercomplex mode[40].

Two-dimensional homonuclear DQF-COSY[41], NOESY[37], and clean-TOCSY spectra[37] were recorded using spectral widths of 6,000 Hz, 2*512 t_1 increments, 1024 complex data points in the acquisition time domain and with 8–32 transients per t_1 increment. NOESY spectra were recorded using cross relaxation mixing times of 60 or 200 ms and clean-TOCSY spectra were recorded using isotropic mixing times of 10, 60 or 80 ms. A 2D ^{1}H,^{13}C-HSQC spectrum was recorded using gradient selection[42] with a ^{1}H and ^{13}C sweep widths of 6000 Hz and 20000 Hz, respectively, 2*128 t_1 increments, 512 complex data points and 160 transients per increment. The HSQC sequence was optimized for a C–H scalar coupling constant of 140 Hz, with the ^{13}C transmitter placed at 57 p.p.m.. 2D SS-NOESY spectra[37] were recorded with a sweep width of 8000 Hz and a 200 ms mixing time. The third pulse in the SS-NOESY sequence is a shifted laminar pulse[36] creating a zero net excitation at the frequency of the transmitter (water resonance). Water suppression was achieved by presaturation of the water signal or presaturation in combination with SCUBA water suppression[43]. No presaturation was used in the HSQC and SS-NOESY experiments.

NMR spectra were processed using software from Varian (VNMR) and/or Biosym Technologies (Felix 2.2). Data processing typically involved apodization with shifted Gaussian functions in the t_2 (acquisition time) domain and *sine/cosine* bell functions in t_1, and

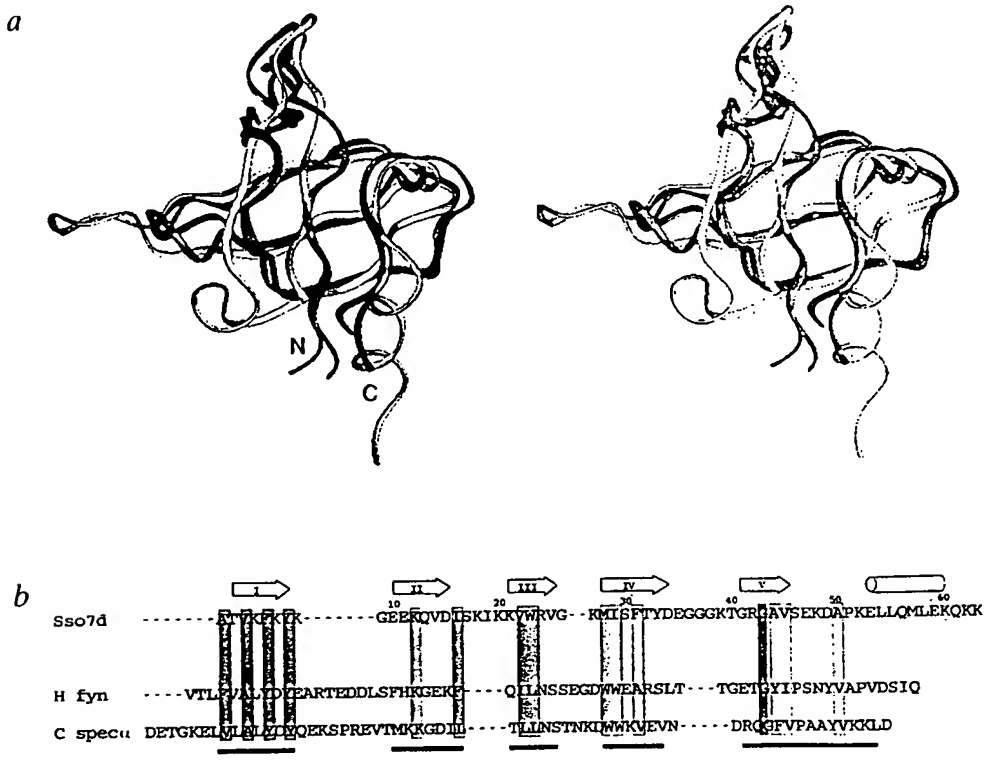


Fig. 7 *a*. Comparison of folding topologies in Sso7d and SH3 domains. The stereo picture contains the superimposed backbones of Sso7d (grey), the SH3 domains of chicken brain α spectrin (green), and the human *fyn* proto-oncogene (blue). *b*, Secondary structure based alignment of the Sso7d sequence to those of the SH3 domains of chicken brain α spectrin (C spec α), and the human *fyn* proto-oncogene (H fyn). Elements of secondary structure in Sso7d are shown at the top. The numbering refers to the Sso7d sequence. The grey bars indicate fragments used in the structure-based alignment. Orange boxes indicate similar or identical hydrophobic residues within the aligned sequences. The blue and green boxes denote a lysine and a glycine which is located at identical positions in the aligned sequences.

baseline correction using routines available within the two software packages. Processed spectra typically contained 1024x1024 real data points.

NMR data analysis. Spin system identification and sequential resonance assignments of ^{1}H resonances in Sso7d were carried out in homonuclear 2D spectra using standard methodologies[18,19]. The natural abundance ^{13}C/^{1}H HSQC spectrum aided significantly when sorting out ^{1}H methyl and aromatic resonances. Most assignment work and collection of NOE constraints were carried out on spectra recorded at 303 K. Analysis of NMR spectra and compilation of NOE data were performed using the interactive computer graphics program ANSIG[39].

Stereospecific assignments of prochiral methylene groups were carried out by identifying predominant χ^1 rotameric states using $^3J_{H_\alpha H_\beta}$ coupling constants measured in DQF-COSY spectra and intraresidue NOEs measured with a short (60 ms) mixing time[35]. The relative magnitudes of $^3J_{H_\alpha H_{\beta 1}}$ and $^3J_{H_\alpha H_{\beta 2}}$ coupling constants could also be measured in clean-TOCSY spectra recorded with a short (10 ms) mixing time using reported simulations[40] as a reference for expected cross peak intensities. Valine methyl groups were stereospecifically assigned and χ^1 rotamers from the magnitude of the $^3J_{H_\alpha H_\beta}$ coupling and the relative intensities of intraresidue $d_{\alpha_i \gamma_i}$ NOE connectivities[41],(note that the notations of valine γ1 and γ2 methyls in ref. 41 are exchanged compared to convention).

The χ^1 rotameric states of Thr 2 and Thr 32 were estimated as follows. Both residues have relatively small $^3J_{H_\alpha H_\beta}$ coupling constants and the HN-Hα cross peaks in DQF-COSY are quadratic[35], indicating that χ^1=60 or χ^1=180. Inspection of the short mixing time NOESY spectrum revealed that $d_{H_\alpha H_\beta} > d_{H_\alpha H_\gamma}$ in Thr 2, which is consistent with χ^1=180, whereas $d_{H_\alpha H_\gamma} > d_{H_\alpha H_\beta}$ in Thr 32, which is consistent with χ^1=60.

NOEs were quantified as distance constraints based on cross peak volumes measured in a NOESY spectrum recorded with a mixing time of 60 ms. The conversion of volumes into distances was based on calibration of observed intraresidue and sequential NOEs within well-defined segments of anti-parallel β-sheet[18]. NOE volumes involving HN protons were corrected for the presence of 10% D_2O in the sample. Cross peak volumes involving methyl protons were divided by three prior to conversion into distance constraints. Distance constraints were divided into four classes: strong (<2.7 Å), medium (<3.3 Å), weak (<5.0 Å) and very weak (<6.0 Å). Pseudoatoms with appropriate distance corrections were created for non-stereospecifically assigned methylene protons, aromatic ring protons and the methyl groups in leucines[18]. A (reduced) pseudoatom correction of 0.3 Å was used to account for effects due to rapid rotation of methyl groups[42].

A total of 14 hydrogen-bonded amide protons could be identified either as slowly exchanging resonances in a TOCSY spectrum of Sso7d dissolved in D_2O, or as amide-proton resonances for which the temperature dependence of the chemical shift is small (< 5 p.p.b.K^{-1}) compared to that of C-terminal residues which are exposed to the solvent (> 8 p.p.b.K^{-1}). These experimentally supported hydrogen bonds (between backbone amide protons and carbonyl oxygens) were imposed in the structure calculations as 28 distance constraints with lower and upper bounds of 1.8 Å and 2.4 Å for amide hydrogen to carbonyl oxygen distances, and 2.6 Å and 3.4 Å for amide nitrogen to carbonyl oxygen distances, respectively. The hydrogen bond constraints were imposed at a late stage of the structure refinement at which point hydrogen bond donor-acceptor pairs could be unambigously identified. All hydrogen bonds used in the calculations are within well-defined regions of anti-parallel β-sheet. A table of sequential assignments of the Sso7d ^{1}H NMR spectrum at 30°C and pH 6.0 is available from the authors on request.

Structure Calculations. Three-dimensional structures were determined using a dynamic simulated annealing (SA) method[43] implemented within the X PLOR 3.0 program[45]. The protocol of Nilges *et al.*[44] was used with some modifications, as described below. Extended peptide conformations were used as starting structures in the simulations. The X PLOR force field—containing potentials for chemical bonds, repulsive van der Waals' interactions and experimental (distance and dihedral) constraints—was used. The k_i constant of the distance constraint potential was set to 50 kcal mol^{-1}Å^{-2} and the force constant of the dihedral (χ^1) square well potential was set to 200 kcal mol^{-1} rad^{-2}. Force constants for planarity and chirality were set to 50 kcal mol^{-1} rad^{-2}. The simulations were carried out in five stages: *i*, 100 steps Powell energy minimization to remove bad non-bonded contacts; *ii*, 15 ps of dynamics at 1000 K with normal van der Waals radii and a low repulsive force constant (0.002 kcal mol^{-1} Å^{-4}); *iii*. 10 ps of 1000 K dynamics during which the repulsive force constant was increased to 0.1 kcal mol^{-1} A^{-4} and the assymptote in the NOE soft square well potential (constant c in ref. 44) was increased from 0.0 to 1.0 (in 10 steps); *iv*, cooling to 300K during 5.6 ps (28 steps of 0.2 ps with 25K cooling/step) with repulsive force constant of 4.0 kcal mol^{-1} A^{-4} and van der Waals' radii scaled by 0.8; and *v*, 1200 steps of Powell minimization with normal van der Waals' radii and force constants for planarity and chirality set to 500 kcal mol^{-1} rad^{-2}. A 1 fs time step was used throughout with bonds constrained using the SHAKE algorithm during stages *i*'–*iv*.

An ensemble of structures was initially calculated after the sequential assignments were almost completed and about 300 distance constraints had been collected. The simulations were then repeated several times during structure refinement. The final round of SA contained 50 simulations out of which 35 converged yielding low energy structures. An average SA structure (SA_{av}) was calculated from the 35 SA structures by averaging superimposed coordinates. The average structure was also minimized (SA_{av_r}) using the same potential as in stage *v* of the SA protocol. The structures were analyzed with respect to the precision of atomic positions and dihedral angles, constraint violations, deviations from idealized bond geometries and non-bonded interaction potentials, and further characterized with respect to dihedral angle conformations and hydrogen bonding. Dihedral angle order parameters, $S^{cos\phi}$, reflecting the precision of the corresponding dihedral within the ensemble were calculated according to Hyberts *et al.*[46]. A value of $S^{cos\phi}$ approaching unity indicates a very well-defined dihedral angle whereas an isotropic distribution yields $S^{cos\phi}$=0 (but $S^{cos\phi}$=0 must not necessarily reflect an isotropic distribution). The ensemble of SA structures were also searched for additional intermolecular hydrogen bonds using the following two criteria: the distance between the donor hydrogen and acceptor oxygen and the two heavy atoms must be less than 2.5 Å and 3.5 Å, respectively. Hydrogen bonds mentioned in the text fulfil these criteria in at least 18 of the 35 SA structures. Structural r.m.s. differences quoted in the text refer to comparisons with the average structure (SA_{av}). It should be noted that r.m.s. difference comparisons containing 'all atoms' can sometimes be erroneous and too large due to the specific atom labelling of phenyl and tyrosyl rings and carboxylate groups. This is because the computer program evaluating r.m.s. differences does not always consider the inherent symmetry of these groups and therefore can give a large r.m.s. difference even in the case of perfect overlap (P. Kraulis, personal communication).

Received 17 August; accepted 16 September 1994.

Acknowledgements This work has been supported by grants from the Swedish Natural Science Council and the Magn. Bergwall Foundation, and by the European Community (project "Biotechnology of Extremophiles") by a grant from NUTEK. We acknowledge the Swedish NMR Centre for providing access to instruments and for assistance. We thank A. Johansson at Pharmacia Bioscience Center for assistance with mass spectrometry.

1. Kornberg, R. & Lorch, Y. Chromatin structure and transcription. *A. Rev. Cell. Biol.* **8**, 563–587 (1992).
2. Schmid, M.B. More than just "histone-like" proteins. *Cell* **63**, 451-453 (1990).
3. Dijk, J. & Reinhardt, R. The structure of DNA-binding proteins from eu and archaebacteria. In *Bacterial chromatin* (Eds, Gualerzi, C.O. and Pon, C.L.) 185–218, (Springer-Verlag, Berlin, 1986).
4. Pettijohn, D. Histone-like proteins and bacterial chromosome structure. *J. biol. Chem.* **236**, 12793–12796 (1988).
5. Drlica, K. & Rouviere-Yaniv, J. Histonelike proteins in bacteria. *Microbiol. Rev.* **51**, 301–319 (1987).
6. Grayling, R.A., Sandman, K. & Reeve, J.N. Archaeal DNA binding proteins and chromosome structure. *System. appl. Micobiol.* **16**, 582-590 (1994).
7. Brock, T.D., Brock, K.M., Belly, R.T. & Weiss, R.L. *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* **84**, 54–68 (1972).
8. Thomm, M., Stetter, K.O. & Zillig, W. Histone-like proteins in eu- and archaebacteria. *Zentralbl. Bakteriol. Mikrobiol. Hyg., I. Abt Orig. C* **3**, 128-139 (1982).
9. Kimura, M., Kimura, J., Davie, P., Reinhardt, R. & Dijk, J. The amino acid sequence of a small DNA binding protein from the archaebacterium Sulfolobus solfataricus. *FEBS Letts.* **176**, 176-178 (1984).
10. Grote, M., Dijk, J. & Reinhardt, R. Ribosomal and DNA binding proteins of the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. *Biochim. biophys. Acta* **873**, 405-413 (1986).
11. Choli, T., Wittmann-Liebold, B. & Reinhardt, R. Microsequence analyses of DNA-binding proteins 7a, 7b and 7e from the archaebacterium *Sulfolobus acidocaldarius*. *J. biol. Chem.* **263**, 7087-7093 (1988).
12. Choli, T., Henning, P., Wittmann-Liebold, B & Reinhardt, R. Isolation, characterization and microsequence analysis of a small basic methylated DNA-binding protein from the archaebacterium *Sulfolobus solfataricus*. *Biochim. biophys. Acta* **950**, 193-203 (1988).
13. Saraste, M., Sibbald, P.R. & Wittinghofer, A. The P-loop - a common motif in ATP- and GTP-binding proteins. *Trends biochem. Sci* **15**, 430-434 (1990).
14. Guaghardi, A., Cerchia, L., Camardella, L., Rossi, M. & Bartolucci, S. DBF (disulfide bond forming) enzyme from the hyperthermophilic archaebacterium Sulfolobus solfataricus behaves like a molecular chaperone. *Biocatalysis* (in the press).
15. Fusi, P., Tedeschi, G., Aliverti, A., Ronchi, S., Tortora, P. & Guerritore, P. Ribonucleases from the extreme thermophilic archaebacterium *S. solfataricus*. *Eur. J. Biochem.* **211**, 305–310 (1993).
16. Jentoft, N. & Dearborn, D. Protein labeling by reductive alkylation. *Meths. Enzymol.* **91**, 570–579 (1983).
17. Stein, D.B. & Searcy, D.G. Physiologically important stabilization of DNA by a prokaryotic histone-like protein. *Science* **202**, 219–221 (1978).
18. Wüthrich, K. *NMR of proteins and nucleic acids*. (Wiley, New York, 1986)
19. Bax, A. Two-dimensional NMR and protein structure. *A. Rev. Biochem.* **58**, 223–256 (1989).
20. Pawson, T & Schlessinger, J. SH2 and SH3 domains. *Curr. Biol.* **3**, 434–442 (1993).
21. Kuriyan, J. & Cowburn, D. Structures of SH2 and SH3 domains. *Curr. Opin. struct. Biol.* **3**, 828–837 (1993).
22. Koyama, S. et al. Structure of the PI3K SH3 domain and analysis of the SH3 family. *Cell* **72**, 945–952 (1993).
23. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. Crystal structure of a src-homology 3 (SH3) domain. *Nature* **359**, 851–855 (1992).
24. Noble, M.E.M., Musacchio, A., Saraste, M., Courtneidge, S.A. & Wierenga, R.K. Crystal structure of the SH3 domain in human *Fyn*; comparison of the three-dimansional structures of SH3 domains in tyrosine kinases and spectrin. *EMBO J.* **12**, 2617–2624 (1993).
25. Falzone, C.J., Kao, Y.-H., Zhao, J., Bryant, D.A. & Lecomte, J.T.J. Three-dimensional solution structure of PsaE from the cyanobacterium *Synechococcus sp.* strain PCC 7002, a photosystem I protein that shows structural homology with SH3 domains. *Biochemistry* **33**, 6052–6062 (1994).
26. Wilson, K.P., Shewchuk, L.M., Brennan, R.G., Otsuka, J. & Matthews, B.W. Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc. natn. Acad. Sci. U.S.A.* **89**, 9257–9261 (1992).
27. Zillig, W. et al. The *sulfolobus*-"Caldariella" group: taxonomy on the basis of the structure of DNA-dependent RNA polymerase. *Arch Microbiol* **125**, 259–269 (1980).
28. Cantor, C.R. & Schimmel, P. *Biophysical chemistry*. Part II, Chapter 7, (Freeman, San Francisco, 1980).
29. Lundbäck, T., Zilliacus, J., Gustafsson, J.-A., Carlstedt-Duke, J. & Härd, T. Thermodynamics of sequence-specific glucocorticoid receptor-DNA interactions. *Biochemistry* **33**, 5955–5965 (1994).
30. States, D.J., Haberkorn, R.A. & Ruben, R.J. A 2D NMR experiment with pure absorption phase in four quadrants. *J. magn. Reson.* **48**, 286–295 (1982).
31. Rance, M. *et al.* Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum coherence. *Biochem biophys. Res. Comm.* **117**, 479–485 (1983).
32. Macura, A. & Ernst, R.R. Elucidation of cross relaxation in liquids by 2D NMR spectroscopy. *Molec. Phys.* **41**, 95–117 (1980).
33. Griesinger, C., Otting, G., Wuthrich, K. & Ernst, R.R. Clean TOCSY for 1H spin system identification in macromolecules. *J. Am. chem. Soc.* **110**, 7870–7872 (1988).
34. Davis, A.L., Keeler, J., Laue, E.D. & Moskau, D. Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients. *J. magn. Reson.* **98**, 207–216 (1992).
35. Smallcombe, S.H. Solvent suppression with symmetrically-shifted pulses. *J. Am. chem. Soc.* **115**, 4776–4785 (1993).
36. Patt, S.L. Single- and multiple-frequency-shifted laminar pulses. *J. magn. Reson.* **96**, 94–102 (1992).
37. Brown, S.C., Weber, P.L. & Mueller, L. Toward complete 1H NMR spectra in proteins. *J. magn. Reson.* **77**, 166–169 (1988).
38. Kraulis, P.J. ANSIG: a program for the assignment of protein 1H 2D NMR spectra by interactive computer graphics. *J. magn. Reson.* **84**, 627–633 (1989).
39. Hyberts, S., Marki, W. & Wagner, G. Stereospecific assignments of side-chain protons and characterization of torsion angles in eglin c. *Eur. J. Biochem.* **164**, 625–635 (1987).
40. Cavanagh, J., Chazin, W. & Rance, M. The time dependence of coherence transfer in homonuclear isotropic mixing experiments. *J. magn. Reson.* **87**, 116–131 (1990).
41. Zuiderweg, E.R.P., Boelens, R. & Kaptein, R. Stereospecific assignments of 1H-NMR methylines and conformation of valyl residues in the lac repressor headpiece. *Biopolymers* **24**, 601–611 (1985).
42. Bartik, K. & Redfield, C. A method for the estimation of χ^1 torsion angles in proteins. *J. biol. NMR* **3**, 415–428 (1993).
43. Koning, T.M.G., Boelens, R. & Kaptein, R. Calculation of the nuclear Overhauser effect and the determination of proton-proton distances in the presence of internal motions. *J. magn. Reson.* **90**, 111–123 (1990).
44. Nilges, M., Gronenborn, A.M., Brünger, A.T. & Clore G.M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine protease inhibitor 2. *Prot. Engng.* **2**, 27–38 (1988).
45. Brünger, A.T. X-PLOR manual. Version 3.0. (Yale University, New Haven, Connecticut, 1992).
46. Hyberts, S.G., Goldberg, M.S., Havel, T.F. & Wagner, G. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Prot. Sci.* **1**, 736–751 (1992).
47. Brooks, B.R. et al. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. comput. Chem.* **4**, 187–217 (1983).

Gene Cloning, Expression, and Characterization of the Sac7 Proteins from the Hyperthermophile *Sulfolobus acidocaldarius*†

James G. McAfee,‡ Stephen P. Edmondson, Prasun K. Datta,§ John W. Shriver,* and Ramesh Gupta*

Department of Medical Biochemistry, School of Medicine, Southern Illinois University, Carbondale, Illinois 62901-4413

Received March 28, 1995; Revised Manuscript Received May 19, 1995®

ABSTRACT: The genes for two Sac7 DNA-binding proteins, Sac7d and Sac7e, from the extremely thermophilic archaeon *Sulfolobus acidocaldarius* have been cloned into *Escherichia coli* and sequenced. The *sac7d* and *sac7e* open reading frames encode 66 amino acid (7608 Da) and 65 amino acid (7469 Da) proteins, respectively. Southern blots indicate that these are the only two Sac7 protein genes in *S. acidocaldarius*, each present as a single copy. Sac7a, b, and c proteins appear to be carboxy-terminal modified Sac7d species. The transcription initiation and termination regions of the *sac7d* and *sac7e* genes have been identified along with the promoter elements. Potential ribosome binding sites have been identified downstream of the initiator codons. The *sac7d* gene has been expressed in *E. coli*, and various physical properties of the recombinant protein have been compared with those of native Sac7. The UV absorbance spectra and extinction coefficients, the fluorescence excitation and emission spectra, the circular dichroism, and the two-dimensional double-quantum filtered ^{1}H NMR spectra of the native and recombinant species are essentially identical, indicating essentially identical local and global folds. The recombinant and native proteins bind and stabilize double-stranded DNA with a site size of 3.5 base pairs and an intrinsic binding constant of 2×10^7 M^{-1} for poly[dGdC]·poly[dGdC] in 0.01 M KH_2PO_4 at pH 7.0. The availability of the recombinant protein permits a direct comparison of the thermal stabilities of the methylated and unmethylated forms of the protein. Differential scanning calorimetry demonstrates that the native protein is extremely thermostable and unfolds reversibly at pH 6.0 with a T_m of approximately 100 °C, while the recombinant protein unfolds at 92.7 °C.

Small basic DNA-binding proteins have been isolated from various archaea, some of which have been shown to be associated with the nucleoid or chromatin and presumably perform a histone-like or helix-stabilizing function in these organisms (Searcy, 1975; Stein & Searcy, 1978; Searcy & Delange, 1980; Thomm et al., 1982; Grote et al., 1986; Lurz et al., 1986; Choli et al., 1988a,b; Reddy & Suryanarayana, 1989; Sandman et al., 1990), although the actual function of many of these proteins has not been demonstrated. HTa protein from the thermophilic archaeon *Thermoplasma acidophilum* shows considerable homology to eukaryotic histones and *Escherichia coli* HU protein (Searcy, 1975; Searcy & Delange, 1980). Hmf1 and Hmf2, two DNA binding proteins from *Methanothermus fervidus*, are also homologous to some of the eukaryotic histones (Sandman et al., 1990).

Sulfolobus, a thermoacidophilic archaeon, expresses a number of small basic DNA-binding proteins ranging in molecular weight from 7000 to 10 000 (Kimura et al., 1984; Grote et al., 1986; Choli et al., 1988a). These have no apparent homology to any of the histones. Much of the early work on these proteins resulted from a search for chromatin proteins that might stabilize the genomic DNA at the high growth temperature. *Sulfolobus acidocaldarius* grows optimally in the range of 70−80 °C, while *Sulfolobus solfataricus* grows optimally at approximately 75−85 °C. The G+C base composition of *Sulfolobus* DNA is about 40%, and its cellular salt concentration is relatively low, making a helix-stabilizing protein presumably necessary (Reddy & Suryanarayana, 1988). The 7 kDa class of proteins has been presented as a likely candidate given that they are present in relatively large amounts in the cell (Grote et al., 1986; Choli et al., 1988a,b).

Five proteins have been isolated in the 7 kDa class from *S. acidocaldarius* (Kimura et al., 1984; Choli et al., 1988b), and have been labeled Sac7a[1] through Sac7e, in order of increasing basicity. Four of these, Sac7a, b, d, and e, have been sequenced (Figure 1) (Kimura et al., 1984; Choli et al., 1988b), and only minor differences among them have been noted. The sequence of Sac7c has not been reported. The number of genes encoding the 7 kDa proteins of *S. acidocaldarius* has not been determined. Comparison of the

† This work was supported by the Biotechnology Research and Development Corporation (J.W.S. and S.P.E.) and the National Institutes of Health (GM49686) (J.W.S. and S.P.E.). A preliminary report of this work was given at the Swedish Biophysical Society Meeting, June 4−6, 1990, Lovanger, Sweden, and at the Biophysical Society Meeting, Feb 9−13, 1992, Houston, TX.

* Authors to whom correspondence should be sent. Phone: 618-453-6479 or 618-453-6466. Fax: 618-453-6440. E-mail: rgupta@som.siu.edu or jshriver@som.siu.edu.

‡ Present address: Vanderbilt University, Department of Molecular Biology, 1161 21st Ave. South, Nashville TN 37235.

§ Present address: Department of Pharmacology and Molecular Biology, Chicago Medical School, N. Chicago, IL 60064-3095.

® Abstract published in *Advance ACS Abstracts*, July 15, 1995.

[1] Abbreviations: DSM, Deutsche Sammlung für Mikroorganismen; IPTG, isopropyl β-D-thiogalactopyranoside; NMR, nuclear magnetic resonance; COSY, correlation spectroscopy; DQF-COSY, double-quantum filtered correlation spectroscopy; DSC, differential scanning calorimetry; CD, circular dichroism; Sac7, a group of 7 kDa DNA-binding proteins from *Sulfolobus acidocaldarius*, individually referred to as Sac7a, Sac7b, Sac7c, Sac7d, and Sac7e, in order of increasing basicity; Sso7, a group of 7 kDa DNA-binding proteins from *Sulfolobus solfataricus*.

0006-2960/95/0434-10063$09.00/0

amino acid sequences indicates that there must be at least two separate genes coding the 7d and 7e species. The high degree of similarity observed in the primary sequence of the 7d and 7e proteins suggests that two genes arose through gene duplication. Sac7a and Sac7b are truncated versions of the Sac7d protein, most likely resulting from truncated genes, posttranslational processing, or degradation during isolation.

Specific ϵ-aminomonomethylation of lysines 4 and 6 is characteristic of the Sac7a, b, and d proteins, while Sac7e is monomethylated at lysines 6, 62, and 63 (residue 4 is an arginine in Sac7e) (Kimura et al., 1984; Choli et al., 1988b). No lysine methylation has been detected in the C-terminus of Sac7a, b, or d, presumably since there are no lysines at positions 62 and 63 in these proteins, although Sac7d contains lysines at positions 64 and 65. The Sso7d protein from *S. solfataricus* is monomethylated at lysines 4 and 6 and also at lysines 62, 64, and 65 (Choli et al., 1988a). The role of lysine monomethylation has not been determined but is most likely nontrivial given the specificity (there are 12–14 lysines in these proteins) and the occurrence in both *S. acidocaldarius* and *S. solfataricus* proteins. Baumann et al. (1994) have recently shown that an increase in Sso7d methylation occurs upon heat shock and indicate that methylation may be directly related to protein stability. However, methylation may be an incidental response to an increase in methylase activity directed at other processes. Methylation may also increase the reversibility of the unfolding process rather than changing the stability. A direct calorimetric measurement of the unfolding and stability of these proteins has not been reported.

The Sac7 proteins would appear to be ideal models for studies of protein folding and stability given their small size, the absence of cysteine, and expected high thermostability. Biophysical analyses of these proteins is hampered, however, by the inability to selectively isolate a homogenous isoform in large quantities. The differential methylation of individual 7 kDa proteins could further complicate quantitative studies of structure and stability as well as DNA binding. Therefore, we have cloned and expressed the gene encoding the Sac7d species in *E. coli* to facilitate elucidation of the solution structure of the protein by NMR with high resolution, probing of the thermostability and DNA-binding properties of the protein by site-directed mutagenesis, and determination of the role of methylation. The availability of recombinant protein allows for a direct comparison of the stability of the methylated and unmethylated proteins. In the process of cloning the *sac7d* gene, the gene for Sac7e has also been cloned and sequenced; and we have delineated the transcription initiation and termination regions of the *sac7d* and *sac7e* genes along with the promoter elements.

An initial structure of the native Sso7d protein has been recently published by Baumann et al. (1994), and a high-resolution structure of the homogeneous, recombinant Sac7d protein has been completed (Edmondson, Qiu, and Shriver, manuscript submitted). There are significant differences between these structures, and it remains to be determined if these can be attributed to sequence differences, lysine methylation, or quality of data due to heterogeneity in the native preparation. The spectroscopic, DNA binding, and calorimetric comparisons of the native and recombinant Sac7 proteins reported here indicate little difference in structure, but significant difference in thermostability.

MATERIALS AND METHODS

Strains of Microorganisms. *E. coli* strain DH5αF'IQ [F' *lacI*^{q}ZΔM15/Δ (*lacZYA-argF*) *recA*1 *hsdR*17(r_k^- m_k^+)] was purchased from Gibco BRL. *E. coli* strains HMS174 (F$^-$ *recA* r^-_{k12} m^+_{k12} Rif^r), BL21 (F$^-$ *ompT* r^-_B m^-_B), and their derivatives were generous gifts from F. William Studier (Studier et al., 1990). *E. coli* strain CJ236 (*dut*$^-$ *ung*$^-$) was obtained from Jack Parker (Southern Illinois University, Carbondale, IL). *S. solfataricus* P2 and *S. acidocaldarius* DG6 were gifts from Dennis Grogan (Grogan, 1989, 1991). *S. acidocaldarius* (DSM 639) and *S. solfataricus* P1 (DSM 5354) were purchased from Deutsche Sammlung für Mikroorganismen (DSM).

The *Sulfolobus* strain used here was received from W. Zillig (originally called *S. solfataricus* P1). We have isolated a single colony of our organism on solid medium (Grogan 1989) and have compared the *Hin*dIII, *Eco*RI, and *Sal* restriction fragment patterns of its genomic DNA with two strains of *S. acidocaldarius* (DG6 and DSM639) and two strains of *S. solfataricus* (DSM5354 and P2) according to Grogan (1989). In each case the restriction pattern of our organism is identical to the *S. acidocaldarius* strains and is distinctly different from the *S. solfataricus* strains. This has been further substantiated by Southern analysis of genomic DNA using Sac7 protein gene specific oligonucleotides (see Results). We have designated our laboratory strain as *S. acidocaldarius* RGJM. There has been confusion in the literature regarding the identity of the strains of two *Sulfolobus* species used in various laboratories at different times. Zillig (1993) has recently addressed this issue and tried to clarify the confusion.

Growth of Microorganisms. *E. coli* strains were grown in Luria Bertani media (1% bactotryptone/1% NaCl/0.5% yeast extract) by standard methods (Sambrook et al., 1989). Small scale cultures of *Sulfolobus* (10–200 mL) were grown in Brock's medium (Brock et al., 1972) at 75 °C, supplemented with 0.2% sucrose. Large scale *Sulfolobus* cultures were grown either in 10 L polypropylene carboy at 78 to 8 °C or in a 16 L VirTis glass fermentor at 70–72 °C with vigorous aeration using DeRosa's medium (DeRosa & Gambacorta, 1975) supplemented with 0.1% glucose and 0.1% glutamic acid.

Enzymes and Chemicals. Restriction enzymes, alkaline phosphatase, T4 DNA ligase, T4 DNA polymerase and T4 polynucleotide kinase were purchased from New England Biolabs, Brisco Ltd., BRL, or United States Biochemical Co. [^{32}P]H_3PO_4 and 5'-[α-^{35}S]adenosine thiotriphosphate triethylammonium salt were purchased from ICN Biochemicals Inc. and Amersham Co., respectively. Sequenase version 2.0 DNA sequencing kit was obtained from United States Biochemical Co. Specific deoxyoligonucleotides were purchased from Research Genetics. The list of the oligonucleotides used in this work is presented in Table 1. Difco bacterial media were purchased from Fisher Scientific. CM52 was obtained from Whatman and Sephacryl S-100 HR from Sigma Chemical Co. All other chemicals were reagent grade and obtained primarily from Fisher Scientific, J. T. Baker Co., and Sigma Chemical Co. Laboratory water was routinely purified to 18.3 MΩ resistance with a recycled Barnstead Nanopure system.

Genomic DNA Isolation. Cells from 10–20 ml cultures were pelleted and resuspended in 0.2–0.3 mL of 10 m

Tris-HCl, pH 8.0/1 mM EDTA/1% SDS. This solution was extracted once each with equal volumes of phenol, phenol/chloroform/isoamylalcohol (25:24:1), and chloroform/isoamyl alcohol (24:1). Sodium acetate (3 M, pH 5.2) was added to the final aqueous phase to a concentration of 0.3 M, followed by DNA precipitation with three volumes of ice-cold ethanol. The DNA was spooled onto a thin glass rod, washed in 70% ethanol, and air dried. The DNA was dissolved in 10 mM Tris-HCl, pH 8.0/1 mM EDTA.

Cloning, Hybridization, and Sequencing. The preparation of a *PstI* genomic library of *S. acidocaldarius* RGJM in *E. coli* strain DH5αF'IQ and screening of the library by colony hybridization was according to published procedures (Berger & Kimmel, 1987; Sambrook et al., 1989). Southern and dot blot hybridizations were carried out using nitrocellulose membranes according to the manufacturer's protocols (Schleicher & Schuell) which are based on the method of Southern (1975). The preparation of [γ-^{32}P]ATP and 5′ ^{32}P-end-labeling of oligonucleotides was by standard methods (Johnson & Walseth, 1979; Gupta, 1984; Sambrook et al., 1989). DNA was sequenced by the dideoxy chain termination method (Sanger et al., 1977) using a Sequenase version 2.0 kit. The final sequences were determined from both strands. The standard universal primers for Stratagene's pBluescript vectors (Short et al., 1988) and specifically synthesized oligonucleotides were used in sequencing reactions. DNA sequences were analyzed using the computer program DNA Inspector IIe (Textco Co.).

Primer Extension. Total RNA from *S. acidocaldarius* RGJM was isolated by previously published procedures (Emory & Belasco, 1990). The primer extension assay was conducted as described in the Promega "Protocol and Applications" manual.

Oligonucleotide-Directed Mutagenesis. Procedures for the oligonucleotide directed mutagenesis were those outlined in the Bio-Rad Muta-Gene manual and are based on Kunkel's method (Kunkel et al., 1987) using *E. coli dut⁻ung⁻* strains. We were unable to propagate the substrate for oligonucleotide directed mutagenesis, pBluescript KS+/*sac7d* (see Results for the description and nomenclature of the plasmids), in *E. coli* strain CJ236 (*dut⁻ ung⁻*). Therefore, we used DH5αF'IQ as the host cell for the production of single-stranded template and as the recipient for transformation with mutagenized plasmid and modified the procedure for the selection of mutant plasmid. Colonies arising from transformation with the plasmids from the mutagenesis reaction to create the *NdeI* site were pooled and grown as a mixed culture. Plasmids isolated from these cells were digested with *NdeI* and separated on a 0.8% agarose gel. Linear plasmids were isolated from the gel, recircularized, and again used to transform DH5αF'IQ. Plasmids were then extracted from individual colonies and screened for the presence of an *NdeI* restriction site by digestion with the enzyme. Final confirmation of the desired mutation in the plasmids was obtained by sequencing.

Gene Expression. For gene expression, pET-3b/*sac7d* was transformed into *E. coli* strain BL21 (DE3) pLysS (Studier et al., 1990). For protein isolation, a 10 mL culture of this transformant was grown overnight in LB broth containing ampicillin (200 μg/mL) and chloramphenicol (27 μg/mL). From this, 0.6−1 mL was used to inoculate 50 mL of fresh medium. At an A_{600} of 0.3−0.6, 25 mL of the culture was diluted into 1 L of new medium. The culture was induced upon reaching an A_{600} of 0.8−0.95 by adding IPTG to a final concentration of 0.4 mM. A small aliquot of each culture was taken prior to induction to assay for expression and plasmid stability as described by Studier et al. (1990). Cultures were harvested at 1 h postinduction and stored at −70 °C.

Protein Isolation and Purification. E. coli cells containing recombinant protein were thawed slowly and resuspended in 100 mL of 10 mM Tris-HCl, pH 7.5/0.5 mM phenylmethanesulfonyl fluoride, and the cells were lysed by repeated freezing and thawing along with brief sonication on ice. To isolate native protein, *Sulfolobus* cells were suspended in 0.05 M KH_2PO_4 buffer (pH 6.8) and lysed by sonication on ice. DNase I (20 mg/100 mL) was added to lysed cells, and the suspension was incubated at 37 °C for 5 min followed by centrifugation at 280000*g* for 60 min. The supernatant was cooled on ice and dialyzed in SpectraPor CE 1000 MWCO tubing against 0.2 M H_2SO_4 overnight at 4 °C. The resulting precipitate was removed by centrifugation at 180000*g* for 30 min, and the supernatant was dialyzed four times against 20 mM Tris-HCl, pH 7.4/1 mM EDTA. A small amount of precipitate was removed by centrifugation, and the supernatant was applied to a CM-52 ion exchange column equilibrated with 20 mM Tris-HCl (pH 7.4). The protein was eluted with a linear NaCl gradient (0.0−0.3 M) with both the native and recombinant Sac7 proteins giving a primary peak at approximately 0.2 M NaCl. Further purification was accomplished by gel exclusion chromatography on Sephacryl S-100-HR in 0.02 M Tris-HCl (pH 7.4).

The identity and purity of the 7 kDa proteins were monitored by nonreducing SDS gel electrophoresis (Schägger & von Jagow, 1987). The recombinant protein showed a single band that comigrated with the mixture of Sac7 native proteins isolated from *S. acidocaldarius* (Figure 2) and was absent in preparations from control *E. coli* cells lacking the recombinant plasmid (data not shown). The Sso7 proteins run slightly ahead of Sac7 proteins, consistent with a molecular weight of 7020 (calculated from the sequence). The Schägger−von Jagow gel used here did not resolve the individual Sac7 and Sso7 native species. The identity of the recombinant Sac7d protein was confirmed by comparison of the double-quantum filtered COSY spectra of native Sac7 and recombinant Sac7d proteins (see below) and by the consistency of the sequence specific ^{1}H NMR assignments with the expected sequence (Edmondson, Qiu, and Shriver, submitted).

In earlier studies the recombinant protein was isolated by a different procedure (McAfee, 1993). *E. coli* cells were lysed and DNase treated as above but without sonication. The pH of the supernatant was adjusted to 1.5 with 5 M H_2SO_4. After 45 min on ice and centrifugation, the supernatant was neutralized with 10 N NaOH. The mixture was incubated in a water bath at 70 °C for 2 h, followed by centrifugation. The supernatant was dialyzed three times with 1 mM NaH_2PO_4 buffer (pH 7.0) followed by CM-52 chromatography as above.

Molecular Weight Determination. Approximate molecular weights of the native and recombinant Sac7 proteins were determined by gel exclusion chromatography on Sephacryl S-100-HR. Cytochrome *c*, myoglobin, carbonic anhydrase, and bovine serum albumin were used as molecular weight standards, and blue dextran and DNP-alanine were used to measure the column void and total volumes, respectively.

The molecular weights were determined as described by Mayes (1984).

Phosphorylation and Glycosylation Assays. Phosphate analysis was performed by the method of Fiske and Subbarow (Fiske & Subbarow, 1925; Leloir & Cardini, 1957). Small aliquots of Sac7 (0.95 mL of a 0.5 mg/mL solution in 0.02 M Tris-HCl, pH 7.0) were incubated at 37 °C for 1 h with 0.05 mL of bovine intestinal alkaline phosphatase (2.5 mg/mL in 0.01 M Tris-HCl, pH 9.8). The protein was precipitated with 0.10 mL of concentrated perchloric acid, incubated on ice for 10 min, and centrifuged for 5 min at 13 000 rpm. To 0.90 mL of supernatant was added 2.0 mL of distilled water, 1.0 mL of 5 N H_2SO_4, 1.0 mL of 2.5% ammonium molybdate, and 0.10 mL of reducing agent [prepared fresh by dissolving 0.25 g of reducing mixture (sodium bisulfite, sodium sulfite, and 1-amino-2-naphthol-4-sulfonic acid in a 46:46:8 ratio) in 10 mL of water]. The solutions were allowed to stand for 20 min, and the absorbance was measured at 660 nm. A standard curve was prepared using known amounts of a 0.01 M KH_2PO_4 solution. *O*-Phosphoserine, treated with alkaline phosphatase as described for Sac7 gave quantitative recovery of phosphate.

The phenol–sulfuric acid reaction was used to assay carbohydrate content of Sac7 protein (Debois et al., 1956; Hirs, 1967). To 1.0 mL aliquots of Sac7 protein solution (0.3 mg/mL) was added 0.25 mL of 80% phenol and 2.5 mL of concentrated sulfuric acid. After mixing, the solutions were left at room temperature for 10 min and then placed in a 25 °C water bath for 20 min. The absorbance was measured at 489 nm. Known amounts of α-D-glucose were used to construct a standard curve.

Protein Extinction Coefficient. Ultraviolet and visible spectra were recorded on a Cary 210 spectrophotometer at 25 °C. The wavelength accuracy was checked using benzene vapor and found to be accurate to within ±0.3 nm, and the absorbsance accuracy was checked using potassium chromate in 0.05 M KOH (Gordon & Ford, 1972) and found to be accurate to within 1%.

The extinction coefficients of both the native Sac7 and recombinant Sac7d proteins were determined by measuring the amino acid concentration using the ninhydrin reaction (Moore & Stein, 1954) for a sample of known absorbance. A standard curve was prepared using amino acid standard H (Pierce Biochemicals) and converted into leucine molar equivalents. The concentration of amino acid standards was checked using tyrosine with an extinction coefficient of $\epsilon_{274.5}$ = 1340 in 0.1 M HCl. The molar concentration of amino acid residues in the samples was calculated by dividing leucine equivalents by the average color yield based on the amino acid composition (Moore & Stein, 1954). The average color yields for Sac7d, lysozyme, and RNase A were 1.0, 1.05, and 1.06, respectively. The extinction coefficients of lysozyme and RNase A standards were checked by this procedure and found to be within 1% of published values. The procedure gave an extinction coefficient of 1.03 ± 0.05 mL/(mg·cm) for both native and recombinant proteins.

The extinction coefficients were also determined by the method of van Iersel et al. (1985) immediately following chromatography of the proteins on Sephadex G-50 in 0.01 M NaH_2PO_4 buffer (pH 6.5). A flat (±0.0005 absorbance units) spectrophotometer baseline was programmed using the same buffer which had been used to equilibrate the column. Protein spectra were collected on samples directly from the gel exclusion column, generally using only those sa with an absorbance less than 2.0 at 205 nm to minimi effects of stray light. The reproducibility of the A_{2} ratio using different aliquots collected through the p peak as it eluted from the column was found to be order of 99%. The linear relationship between the extin coefficient at 280 nm and the ratio of the absorbance and 205 nm was confirmed in our hands using b α-chymotrypsin (Worthington), hen egg white lysc (Sigma), bovine pancreatic ribonuclease A (Sigma), (Sigma), β-lactoglobulin (Sigma), and bovine serum all (Sigma). A linear fit of the standards yielded a sta curve such that

$$\epsilon_{280}^{0.1\%} = 35.76\frac{A_{280}}{A_{205}} - 0.04$$

with a correlation coefficient of 0.999 and a sta deviation for the slope of 0.62 and 0.03 for the y inte The extinction coefficients for the native and recom protein were found to be identical with this technique mL/(mg·cm) with a standard deviation of 0.008 mL/(m

The extinction coefficients were also calculated to b mL/(mg·cm) in 6 M guanidine hydrochloride, based amino acid content of the protein using the proced Edelhoch (Edelhoch, 1967; Gill & von Hippel, assuming ϵ_{Tyr} = 1280 M^{-1} cm^{-1}, ϵ_{Trp} = 5690 M^{-1} c 6 M guanidine hydrochloride. An increase in absor of 3.5% was noted upon denaturation of the protein M GdnHCl, so the calculated extinction coefficient folded protein was corrected to 1.05 mL/(mg·cm). estimated error was taken to be ±0.04 with a maxima of ±0.15 (Gill & von Hippel, 1989).

Circular Dichroism. Circular dichroism spectra of p native Sac7 and recombinant Sac7d proteins were me at room temperature in a 0.01 cm path length cylin cell on an AVIV 62DS spectropolarimeter. CD dat collected at 1 nm intervals using averaging times of s/nm, depending on the signal-to-noise ratio. Relativel signal-to-noise ratios made signal averaging of multiple unnecessary. The spectral bandwidth was 1.5 nm. Ba were measured using water and subtracted from the s CD. Sample concentrations ranged from 0.2 to 0.7 m Protein concentrations were determined from UV abso spectra measured in 1 cm cuvettes. The molar C peptide bond was determined using standard proc (Johnson, 1984) along with the UV extinction coef determined above. CD spectra were smoothed as des by Savitsky and Golay (1964). The CD was calibr 290.5 nm with *d*-camphor-10-sulfonic acid using Δ 2.36, and the ratio $\Delta\epsilon_{192.5}/\Delta\epsilon_{290.5}$ was −2.10 (Chen & 1977).

The fractions of protein secondary structures were mined by fitting the CD spectra from 260 to 184 n nm intervals using the variable selection method of J (Manavalan & Johnson, 1987). The results reported averages plus or minus one standard deviation of all p combinations of 22 reference proteins taken 19 at that (1) have secondary structure components great −0.05, (2) have sums of secondary structures betwe and 1.1, and (3) have an rms error between measur calculated CD spectra less than 0.21 Δε units. The n

of fits meeting this selection criteria were greater than 250 for native and recombinant protein.

Nuclear Magnetic Resonance. NMR spectra were collected on a Varian 500 MHz NMR spectrometer with the magnet installed on a TMC Micro-g triangular antivibration table. All data were collected at 35 °C in 90% H_2O/10% D_2O, pH 4.1, with a protein concentration of approximately 10 mM. The pH was adjusted with DCl and NaOD using a Radiometer glass electrode and was not corrected for the deuterium isotope effect (Bundi & Wüthrich, 1979). The chemical shifts are referenced to the water resonance at 4.73 ppm at 35 °C [measured relative to sodium 4,4-dimethyl-4-silapentane sulfonate (DSS) in a separate experiment without protein].

Phase-sensitive double-quantum filtered COSY (DQF-COSY) spectra were collected using standard procedures (Rance et al., 1983). Typically, 1024 data points were collected in the t_2 domain with 512 increments in the t_1 domain, each the sum of 32 scans with a 3 s relaxation delay. The spectral widths in both dimensions was 6000 Hz. The water peak was diminished in all experiments by presaturation during the relaxation delay. Both carrier and decoupler frequencies were set equal to the water resonance frequency in all experiments (Zuiderweg et al., 1986).

The NMR data were transferred to a Silicon Graphics workstation for Fourier transformation and further data manipulation using FELIX 2.1 (BioSym). The data were zero-filled to 2048 data points in both dimensions and treated with a Lorentzian to Gaussian apodization function prior to Fourier transformation.

Differential Scanning Calorimetry. Differential scanning calorimetry was performed with a Microcal MC2 calorimeter. Temperature calibration was monitored using sealed samples supplied by Microcal. Heat flow accuracy was periodically monitored by applying pulses of known magnitude using the internal heater. In addition, ribonuclease A (Sigma, R5250) was used as a benchmark test protein and shown to unfold at pH 2.2 [0.1 M KCl, 0.02 M glycine, ϵ_{280} = 0.69 mL/(mg·cm), MW 13 700] with a T_m of 36.0 °C, a ΔH_{cal} of 74.1 kcal/mol, and a ΔH_{vh} of 74.8 kcal/mol ($\Delta H_{cal}/\Delta H_{vh}$ ratio of 1.00 ± 0.01), in good agreement with the published values of Tiktopulo and Privalov (1974).

Protein solutions were exhaustively dialyzed against the indicated buffer overnight. The sample cell was loaded with 1.229 mL of protein solution, and the reference cell was filled with the last dialysis buffer. Approximately 30 psi of nitrogen was applied to the cells during each scan to minimize degassing during heating. Samples were not degassed, but, instead, the sample was heated repetitively three times in the DSC instrument by scanning to 35 °C (i.e., below any denaturation endotherm), followed by rapid cooling. This procedure resulted in the flattest and most reproducible instrumental baselines.

All DSC experiments were under computer control using an IBM PC computer interfaced to the Microcal MC2 instrument. A scan rate of 1 deg/min was used in all experiments. The computer interface and data collection software were supplied by Microcal. Multiple, repetitive scans were performed on the same sample to check for reversibility, with identical cooling and equilibration times between scans.

The DSC raw data, in the form of heat flow (mcal/min) as a function of temperature, was transferred to a Macintosh Quadra computer for analysis. The raw data were converted to excess heat capacity (kcal/deg·mol) by dividing each data point by the scan rate and the concentration of protein in the sample cell. All baselines were corrected by subtraction of DSC scans of the buffer against which the protein had been dialyzed. The heat capacity data was fit by using in-house nonlinear least-squares fitting routines to obtain the midpoint temperature of the transition and both the calorimetric and van't Hoff enthalpies. The basis of the programs has been described elsewhere (Shriver & Kamath, 1990).

Fluorescence. Fluorescence titration measurements were performed on an SLM 8000C spectrofluorimeter with 4 nm excitation and 8 nm emission slit widths. Binding titrations were performed with excitation at 295 nm and emission monitored at 350 nm. Reverse titrations were performed by adding aliquots of concentrated nucleotide solutions to a known concentration of protein in a 4 mL fluorescence quartz cell with stirring using a magnetic "flea" within the cell. Nucleic acid concentrations were determined spectrophotometrically using an extinction coefficient of 8400 L/(cm·mol) for poly[dGdC]·poly[dGdC] (Wells, 1970) and 6600 L/(cm·mol) for poly[dAdT]·poly[dAdT] (Inman, 1962). All experiments were performed at 25 °C. The fluorescence intensity was constant at high DNA concentrations, and thus no correction was made for the inner filter effect. Apparently, any decrease in fluorescence due to the inner filter effect was balanced by other effects, such as scattering by the DNA-protein complexes. Photobleaching was not observed during the titrations. Binding parameters were obtained by using a simple, noncooperative McGhee–von Hippel model (McGhee & von Hippel, 1974).

DNA Stabilization. Thermal denaturation studies of DNA and DNA–protein complexes were performed on a Cary 210 spectrophotometer equipped with water-jacketed cuvette holders and a circulating water bath calibrated to within ±0.3 °C. Melting curves are scaled to an A_{262} of 1.0 at 20 °C for the DNA component of DNA–protein mixtures.

Sequence Analysis. BLAST (Altshul et al., 1990) searching and alignment were performed using the NCBI server (blast@ncbi.nlm.nih.gov) against the "nr" (nonredundant) sequence database (including Brookhaven Protein Data Bank, January 1994 release; SWISS-PROT Release 29.0, June 1994; PIR Release 41.0, June 30, 1994; CDS Translations from GenBank Release 83.0, June 15, 1994, Kabat Sequences of Proteins of Immunological Interest Release 5.0, August, 1992; TFD Transcription Factor Database Release 7.0, June 1993). BLITZ and FASTA searches of the latest SWISS-PROT database were performed using the EMBL servers (blitz@embl-heidelberg.de and fasta@embl-heidelberg.de). Database retrieval was performed using the GDB/Accessor (Johns Hopkins University) available from ftp.gdb.org. MacPattern (Fuchs, 1991) (fuchs@embl-heidelberg.de) was utilized for BLOCKS (Henikoff & Henikoff, 1991) and PROSITE (Bairoch, 1992) analysis on a Quadra 700 (BLOCKS database Version 7.01 was utilized with 2679 entries and PROSITE database version 12.0, June 1994, was used with 1021 entries, both obtained from the NCBI ftp site ncbi.nlm.nih.gov.) The MacVector software package (IBI) was utilized for protein secondary structure analysis.

Table 1: List of Oligonucleotides

oligonucleotide[a]	sequence[b]	position[c]
A	NACYTCYTTYTCYTCNCC	230−247
B	GGGAGCTTYAARTAYAARGGNGARGA[d]	218−237
C	GGGGTACCRTTRTCRTCRTANGTRAA[d]	296−317
D	TCTTAACAAATTATTTTATTT	398−418
E	GCCCTTTATACCTTCCCCTTA	398−418
F	CCTGTCTTACCATTGTCGTC	305−324
G	CCTTCACCATATGAGGTCAAGTTATC[e]	187−212
H	GACTTAACTTAATACCG	143−159

[a] Oligonucleotides A, B, and C were derived from amino acids 9−14, 5−11, and 31−38, respectively, of the Sac7 proteins (Figure 1). These amino acid sequences are identical in the four Sac7 proteins. [b] N = A, G, C, or T; Y = C or T; R = A or G. [c] Nucleotide positions correspond to those in Figure 3. Sequences of oligonucleotides A, C, D, E, F, and G are complementary to the sequences shown in Figure 3. Oligonucleotides D and E correspond to the same positions (Figure 3) for *sac7d* and *sac7e*, respectively. [d] Oligonucleotides B and C have six and four additional nucleotides, respectively, at the 5′ termini which are not derived from the amino acid sequence of the protein. [e] Sequence of the primer used for oligonucleotide directed mutagenesis. The underlined G replaces a T in the *sac7d* gene sequence creating an *Nde*I restriction site.

RESULTS

Gene Cloning and Sequence. *Pst*I digested genomic DNA of *S. acidocaldarius* RGJM was shotgun cloned in the vector pUC19 and transformed into *E. coli*, DH5αF′IQ. Approximately 10 000 transformants were screened by colony hybridization to a mixed oligonucleotide probe (oligonucleotide A, Table 1) derived from residues 9−14 of the published amino acid sequence of the *S. acidocaldarius* 7 kDa proteins (Kimura et al., 1984; Choli et al., 1988a). [The published amino acid sequences for Sac7a, b, d, and e are identical over this range (Figure 1) as well as over the ranges for oligonucleotides B and C.] Tentative positive clones were restreaked onto selective media and screened a second time with the same probe. Plasmids isolated from a number of these positive clones were then independently hybridized to three different mixed probes (oligonucleotides A, B, and C, Table 1) by dot blot hybridization. Two clones were isolated which hybridized to all three probes. Plasmids isolated from these cells were partially sequenced using oligonucleotide B as a primer. One of the genes corresponded with the published protein sequence for the carboxy-terminal half of the Sac7d protein of *S. acidocaldarius* (Kimura et al., 1984; Choli et al., 1988a) with the exception of one additional lysine at the carboxy terminus, and the other corresponded to the Sac7e sequence. The genes which matched the Sac7d and 7e proteins have been designated *sac7d* and *sac7e*, respectively.

Agarose gel analysis of the plasmids carrying the *sac7d* (pUC19/*sac7d*) and *sac7e* (pUC19/*sac7e*) genes indicated that the cloned *Pst*I fragments were greater than 15 kb in size. Southern blot hybridizations of oligonucleotide C to the restriction digests of pUC19/*sac7d* indicated that *sac7d* gene was present on a slightly less than 800 bp *Eco*RI fragment. Preliminary sequencing of pUC19/*sac7d* using oligonucleotide B as a primer indicated the presence of an *Eco*RI site 61 bases downstream of the termination codon of the protein. Since the published sequence of Sac7d protein consists of 64 amino acids (Kimura et al., 1984; Choli et al., 1988a), the second *Eco*RI site was expected to be upstream of the start codon. Thus, the *Eco*RI fragment hybridizing to probe C was expected to contain the coding region of the gene. This *Eco*RI fragmen subcloned in the vector pBluescript KS+ to produce script KS+/*sac7d*, and the sequence of *sac7d* gen determined (Figure 3). The sequence of the *sac7e* (Figure 3) was obtained directly from the pUC19/*sac7e* primers complementary to the coding region of the g

The GenBank accession numbers for the *sac7d* and gene sequences reported here are M87569 and L respectively.

Sequence Analysis and Gene Copy Number. The s transcription for both *sac7d* and *sac7e* genes was deter using primer extension analysis (Figure 4). Specific p (oligonucleotides D and E, Table 1) that were complem to residues 398−418 (Figure 3) of the two genes were A single start site was observed for each of the two which occurs on a guanosine residue eight nucle upstream from the initiation codon. These gua residues are present within perfect archaeal "B box" sensus sequences (consensus $\frac{A}{T}TG\frac{A}{C}$ (Zillig et al., 19 sequence resembling the archaeal "A-box" motif (con TTTA$\frac{A}{T}$A) is seen 24 and 23 nucleotides upstream fro transcription start site for the *sac7d* and *sac7e* respectively (Figure 3). The "A-box" of *sac7d* has base match with the consensus sequence, while that f *sac7e* has only four matches.

Oligonucleotide F (Table 1) was used to probe ge blots of three *S. acidocaldarius* (RGJM, DG6, and DS and two *S. solfataricus* (DSM5354 and P2) strains (I 5A). Oligonucleotide F is complementary to a region c for residues 34−40 (Figure 1) which are identical for *S. acidocaldarius* 7 kDa proteins (DDNGKTG) and si cantly different from that of *S. solfataricus* (DEGGG two substitutions and an insertion). Two *Hin*dIII resti fragments (~3.0 and ~4.6 kb) were recognized by the in all three *S. acidocaldarius* strains, while no hybridi to the *S. solfataricus* strains was observed. This obser reinforces the assignment of the RGJM strain (our labo strain) as an *S. acidocaldarius* strain. The results in that the putative genes encoding all of the Sac7 protei present on the two *Hin*dIII restriction fragments of ~3. ~4.6 kb in size. Genomic blots of *Eco*RI, *Hin*dIII, an digested *S. acidocaldarius* RGJM DNA were also p with the common oligonucleotide F (Figure 5B), and i case hybridization to two bands was observed. One in each hybridized to oligonucleotide H, specific f untranscribed region upstream of the *sac7d* gene (Figur Results of the hybridizations of various restriction d of the original pUC/*sac7d* and pUC/*sac7e* clones t propriate oligonucleotides (data not shown) corroborat results in Figure 5 and also indicated that the original had a single copy of a *sac7* gene. The 3.0 and 4.6 kb *H* fragments can be correlated with the *sac7d* and *sac7e* respectively. The data indicate that there are only tw genes in *S. acidocaldarius* genome, each being preser single copy. This reinforces the conclusion that Sac7 Sac7b are proteolytically truncated versions of the protein.

Protein Sequence Analysis. The *sac7d* open reading can encode a 66 amino acid protein with a calc molecular weight of 7608, and the *sac7e* encodes a 65 acid protein with a calculated molecular weight of

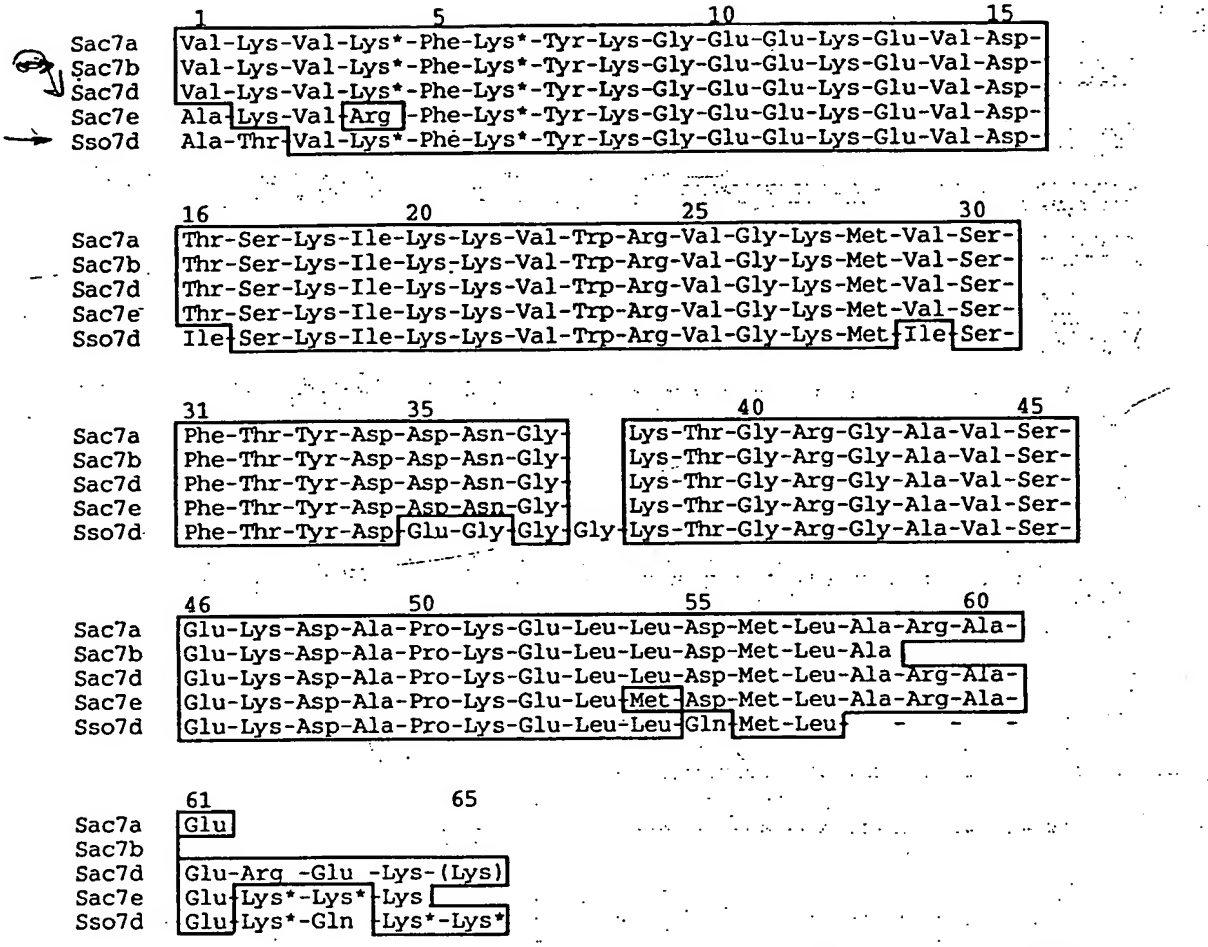


FIGURE 1: Amino acid sequences of the Sac7a, b, d, and e proteins [after Kimura et al. (1984) and Choli et al. (1988b)] and the Sso7d protein [after Choli et al. (1988a)]. [Note that the sequence reported by Kimura et al. (1984) was claimed to be for Sso7d but was later shown to be for Sac7d (Choli et al., 1988a).] Numbering is according to the Sac7d sequence without the initiator methionine. Regions homologous to the Sac7d protein are outlined. Sac7a, b, and d differ only in length. Lysines which are monomethylated to some extent in the native protein are indicated with asterisks. The additional C-terminal lysine coded by the *sac7d* gene described here which was not indicated in the published protein sequence is enclosed in parentheses.

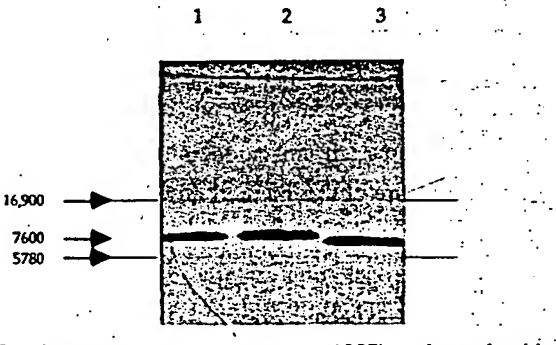


FIGURE 2: Schägger and von Jagow (1987) polyacrylamide nonreducing SDS gel of purified native Sac7 proteins (lane 1), recombinant Sac7d (lane 2), and native Sso7 (lane 3) proteins stained with Coomasie Brilliant Blue G-250 (Bio-Rad). The molecular weight of the Sso7 protein is 7019 based on the published protein sequence (Choli et al., 1988a), while that of the Sac7d is 7608 based on the DNA sequence presented here. The band positions of myoglobin (MW 16 900) and insulin (MW 5780) are indicated for comparison.

(including initiator methionines). Secondary structure analysis of the sequences of the Sac7d and Sac7e proteins was performed with both the Chou–Fasman (Chou & Fasman, 1974, 1978) and the Robson–Garnier algorithms (Robson & Suzuki, 1976; Garnier et al., 1978). Both methods predict the occurrence of significant α-helix (52%) in both proteins extending from approximately Lys9 to Lys28 and from Gly43 to Ala59. Only the Chou–Fasman algorithm predicts a small amount of β-sheet (12%) extending from Lys22 to Lys29 and from Ser31 to Asp36. Reverse turns are predicted near Asp36 and Gly43. These predictions are not consistent with the solution structure of the Sac7d protein which has been determined by 2D NMR (Edmondson, Qiu, and Shriver, manuscript submitted).

Recombinant Gene Expression. The *sac7d* gene (in pBluescript KS+/*sac7d*) was modified by converting the hexanucleotide sequence containing the initiation codon (AATATG) to an *Nde*I site (CATATG) by oligonucleotide G (Table 1) directed mutagenesis to produce pBluescript KS+/*sac7d*(Nd). The *Nde*I–*Bam*HI fragment of pBluescript KS+/*sac7d*(Nd) carrying the coding region of *sac7d* gene was then subcloned into the *Nde*I–*Bam*HI site of pET-3b (Studier et al., 1990) to give pET-3b/*sac7d*, and transformed into HMS174 (DE3), HMS174 (DE3) pLysS, BL21 (DE3), and BL21 (DE3) pLysS (Studier et al., 1990). The plasmid could be established in all of these strains except BL21 (DE3). Furthermore, in transformed BL21 (DE3) pLysS, the growth of the organism is impaired and cultures lyse within 60–70 min after induction with IPTG. On the other hand, the growth of HMS174 strains were not significantly effected by the presence of the plasmid, and lysis was not observed in cultures after 3 h postinduction. The absence of impaired growth in the presence of the plasmid in these

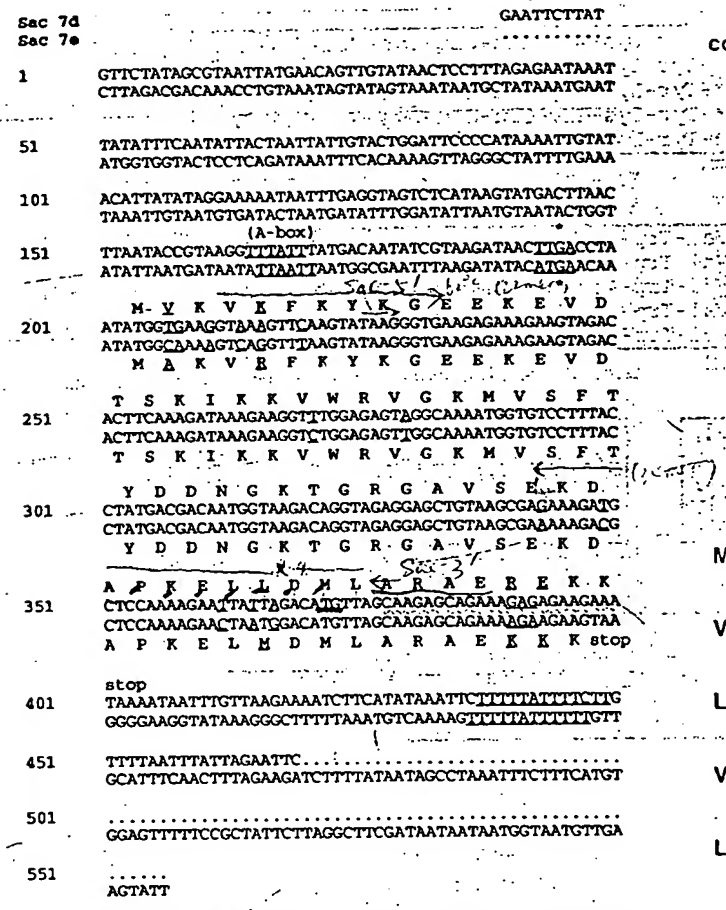


FIGURE 3: Nucleotide sequences of the *sac7d* and *sac7e* genes. The top and bottom sequences are the nucleotide sequence for the *sac7d* and *sac7e* genes, respectively (aligned using the coding region of each gene). Numbering starts with the *sac7e* sequence. The amino acid sequence coded for by each gene is shown above (*sac7d*) or below (*sac7e*) each nucleotide sequence. Putative promoter (A- and B-boxes) and termination elements are underlined in the 5′ and 3′ noncoding regions of each sequence. Amino acid and nucleotide differences in the coding region of each gene are also indicated by underlines. The G at the start of transcription (in the B-box) for each gene is indicated with an asterisk.

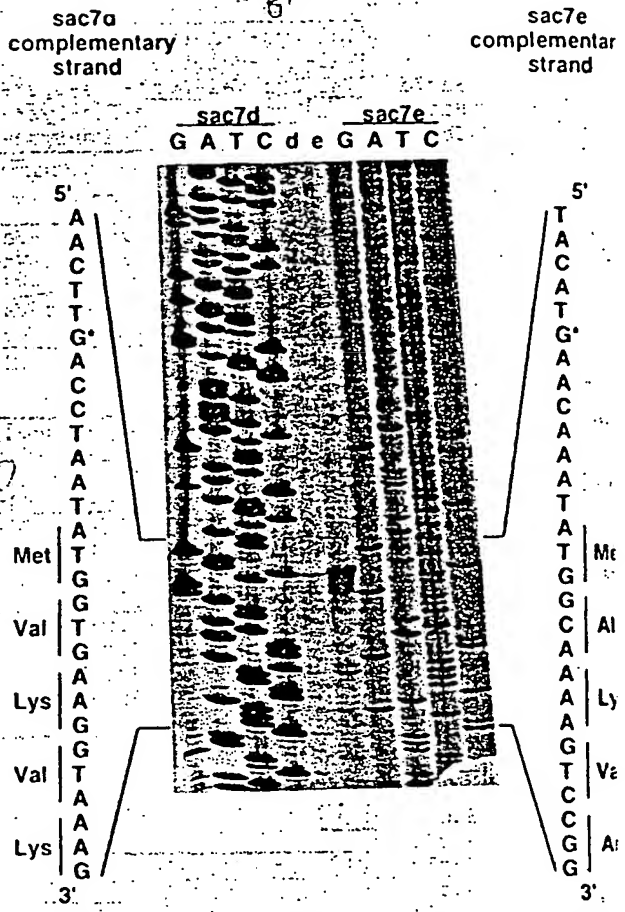


FIGURE 4: Determination of the *in vivo* start of transcription f the *sac7d* and *sac7e* genes by primer extension analysis. *sac7d* (la d) and *sac7e* (lane e) specific oligonucleotides D and E, respectiv [which are complementary to residues 398−418 (Figure 3)], w used to prime the synthesis of a complementary strand of DI from total *S. acidocaldarius* RNA. These same oligonucleoti were also primers in the dideoxy sequencing reactions used markers for the *sac7d* (pBSKS+/*sac7d*) and *sac7e* genes (pUC *sac7e*) indicated. The sequences written on the left and right complementary to the ones observed in the autoradiogram in marked region. The start of transcription is indicated in e sequence by an asterisk. The first five coded amino acids of e protein are also indicated along side each complementary str sequence.

strains was correlated with a lack of Sac7d protein accumulation. In contrast to HMS174 strains, BL21 and its derivatives lack the *ompT* outer membrane protease and are deficient in the *lonA* protease (Studier et al., 1990). The *ompT* protease has been shown to be responsible for T7 RNA polymerase degradation during protein purification from *E. coli* (Grodberg & Dunn, 1988). Thus, it appears that in the absence of stringent regulation of T7 RNA polymerase synthesis prior to induction with IPTG, or proteolytic degradation of the Sac7d protein, the protein accumulates to lethal levels. However, because significant amounts of the Sac7d protein do not accumulate in HMS174 strains, we have utilized BL21 (DE3) pLysS for subsequent expression and purification of the protein.

Spectroscopic and Chemical Characterization. The UV spectra of native and recombinant Sac7 proteins were essentially identical, as expected, given the presence of a single tryptophan and two tyrosines and two phenylalanines in all proteins. The calculated extinction coefficient based on amino acid composition is 1.05 mL/(mg·cm) at 280 nm, in good agreement with the value of 1.03 mL/(mg·cm) determined by ninhydrin analysis. The extinction coefficients were also determined by using the ratio of absorbance at 280 and 205 nm (see Materials and Methods). The empirical nature of this method might lead to some quest of its accuracy, but the high correlation of the results fr the six standards is extraordinary ($r = 0.999$), and reproducibility of the A_{280}/A_{205} ratio measurement is h leading to an expected error of 0.6%. The ratio meth demonstrates that the extinction coefficients of the nat and recombinant protein are identical, viz., the mean of extinction coefficient measurements (native and recombi combined) using this method was 1.18 mL/(mg·cm) wi standard deviation of 0.008 mL/(mg·cm). The final extr tion coefficient for both the recombinant and native prot is taken to be 1.09 mL/(mg·cm), the mean of the t independent measurements, with a standard error of ±((calculated by propagating the errors of the three meas ments). The extinction coefficient was shown to be independent from 2 to 10.

The fluorescence excitation and emission spectra of native Sac7 and recombinant Sac7d proteins were essentially identical (data not shown). In addition, fluorescence emission spectrum was essentially that expe

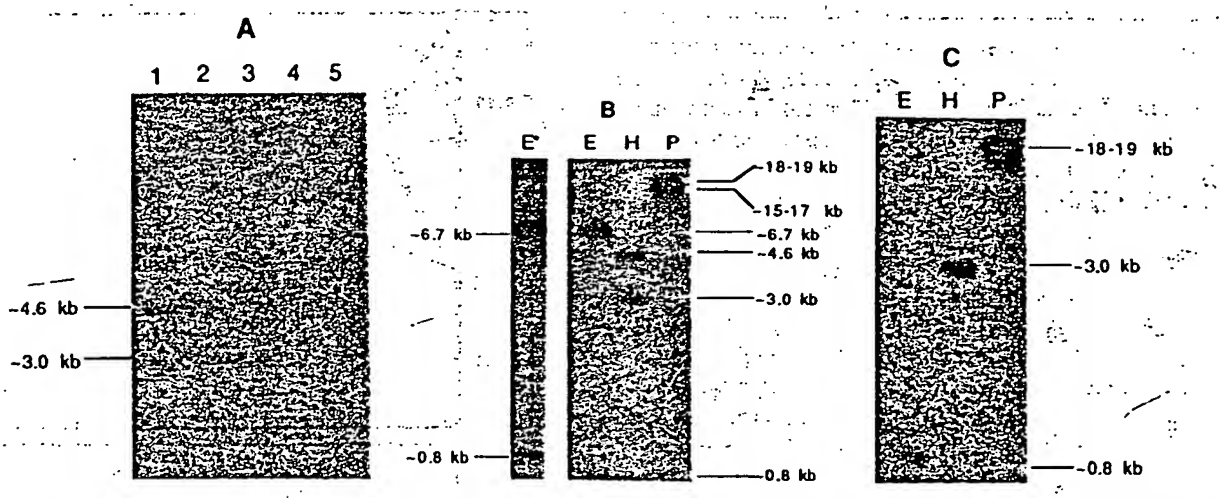


FIGURE 5: Southern analysis of *Sulfolobus* genomic DNA. (A) Autoradiogram of a Southern blot of *Hin*dIII digests of genomic DNA from *S. acidocaldarius* (RGJM) (lane 1), *S. acidocaldarius* (DG6) (lane 2), *S. acidocaldarius* (DSM639) (lane 3), *S. solfataricus* (DSM5354) (lane 4), and *S. solfataricus* (P2) (lane 5) probed with oligonucleotide F. The approximate sizes of the restriction fragments hybridizing to oligonucleotide F are indicated. (B) Autoradiogram of a Southern blot of *Eco*RI (lane E), *Hin*dIII (lane H), and *Pst*I (lane P) digested *S. acidocaldarius* RGJM genomic DNA hybridized with oligonucleotide F. Two closely spaced bands in lane P are clearly evident in the original autoradiogram. Lane E* is a second independent *Eco*RI experiment to clearly demonstrate the 0.8 kb fragment. (C) Similar to panel B except that the DNA was probed with oligonucleotide H.

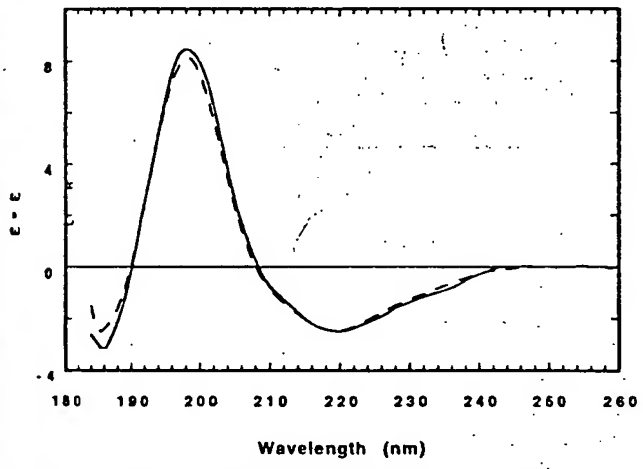


FIGURE 6: Circular dichroism spectra of native Sac7 (solid line, 0.26 mg/mL) and recombinant Sac7d (dashed line, 0.66 mg/mL) proteins in 0.01 M KH_2PO_4, pH 7.0.

for a free tryptophan, indicating that the single tryptophan is highly solvent exposed in both proteins. Notably, the fluorescence emission spectra show a small shift upon DNA binding (data not shown), indicating that the exposure of the tryptophan changes slightly upon DNA binding. The CD spectra of native Sac7 and recombinant Sac7d proteins were also essentially identical (Figure 6). The variable selection method of Johnson (Manavalan & Johnson, 1987) indicates that both the native and recombinant Sac7 proteins are composed of 31% helix (both α- and 3_{10}-helix), 22–25% β-sheet, 0–2% turn, and 42–45% nonrepetitive structure.

The DQF-COSY spectra of the native and recombinant Sac7 proteins are remarkably similar (Figure 7). The native spectrum shows some additional correlation peaks, most likely due to the presence of 7a, b, c, d, and e isoforms in the native preparation and posttranslational modifications (e.g., monomethylation of lysines) in *Sulfolobus*. The essential identity of the chemical shifts for the native and recombinant proteins indicates again that the recombinant and native proteins are folded similarly. The extensive number of alpha protons shifted downfield of the water line at 4.7 ppm indicates the presence of significant β-sheet structure (Wishart et al., 1992). The wide chemical shift dispersion has permitted an essentially complete assignment of the proton resonances and determination of the solution structure (Edmondson, Qiu, and Shriver, manuscript submitted).

No phosphorylation or glycosylation of either the native or recombinant proteins could be detected. The recombinant protein differs from the native by containing the initiator methionine. The recombinant protein also contains an additional C-terminal lysine which was not reported in the amino acid sequence (Kimura et al., 1984), although it remains to be determined if this is an error in the protein sequence or if the lysine is actually removed posttranslationally.

DNA Binding. The binding of Sac7 proteins to DNA is associated with a significant quenching of the intrinsic fluorescence of the single tryptophan (Trp23) in both the native and recombinant Sac7 proteins (Figure 8). Binding of poly[dGdC]·poly[dGdC] in 0.01 M KH_2PO_4 at pH 7.0 leads to a maximal fluorescence quenching of the native protein by 88% and the recombinant Sac7d protein by 87%. Poly[dAdT]·poly[dAdT] shows a maximal quenching of 84% for both proteins (data not shown). The binding data can be fit using the McGhee and von Hippel model (McGhee and von Hippel, 1974) without cooperative interactions assuming a linear relationship between fractional quenching and protein binding. The poly[dGdC]·poly[dGdC] data can be fit with an intrinsic association constant of 2×10^7 M^{-1} for both native and recombinant Sac7d protein and site sizes of 7 bases (3.5 base pairs) and 6.8 bases for native and recombinant protein, respectively. Poly[dAdT]·poly[dAdT] appears to bind slightly weaker with an association constant of 1×10^7 M^{-1} for both proteins and site sizes of 7.5 bases for native protein and 6.8 bases for recombinant protein.

The binding of Sac7 to poly[dAdT]·poly[dAdT] significantly stabilizes the DNA double helix against thermal denaturation. The UV melting curve of poly[dAdT]·poly[dAdT] in 0.01 M KH_2PO_4 is very sharp and has a T_m of 43.5 °C (Figure 9). In the presence of native Sac7d protein,

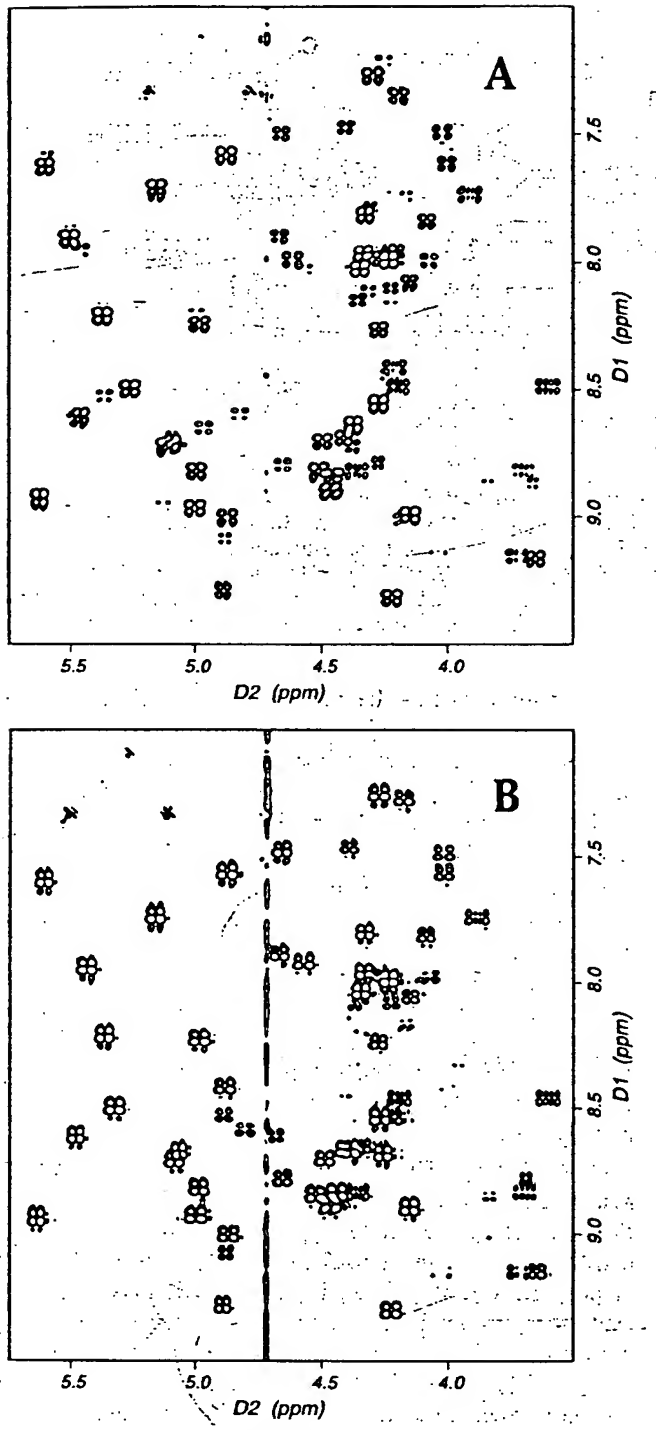


FIGURE 7: Double-quantum filtered (DQF-COSY) α to amide proton correlation spectra of the native Sac7 (A) and recombinant Sac7d (B) proteins at 35 °C in 90% H_2O/10% D_2O, pH 4.1. The protein concentrations in both spectra were approximately 10 mM.

the melting profile of poly[dAdT]·poly[dAdT] broadens and the T_m increases. At the highest protein concentration used in this series of experiments, the DNA melting temperature was increased about 33 °C above that of poly[dAdT]·poly-[dAdT] alone. The recombinant protein increases the T_m of poly[dAdT]·poly[dAdT] by a similar amount. However, the recombinant protein differs in that it aggregates as the double-stranded poly[d(AT)] melts. CD measurements of the suspension, and the supernatant after allowing the aggregate to settle, indicate no major conformational changes during aggregation of the protein–DNA mixture.

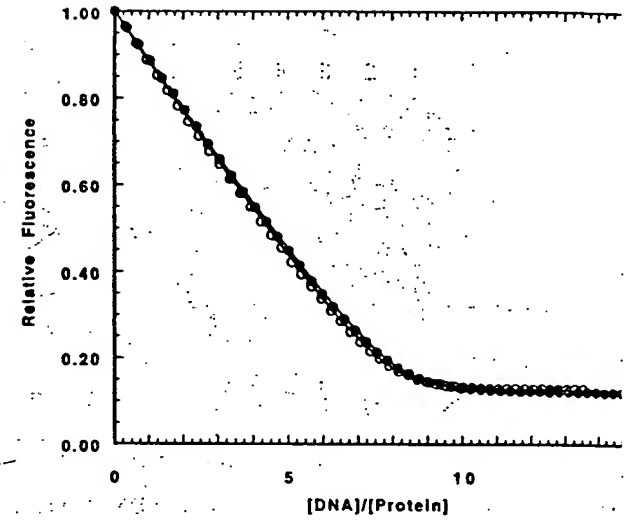


FIGURE 8: Reverse titrations of the native Sac7 (solid circles) a recombinant Sac7d (open circles) proteins with poly[dGdC]·po [dGdC] at pH 7.0 (0.01 M KH_2PO_4), 25 °C with 6.6 μM Sa proteins and 7.3 μM Sac7d. The smooth curves through the d are overlays of simulations using a noncooperative McGhee–v Hippel model (McGhee & von Hippel, 1974). For the native Sa proteins this corresponds to a site size of 7 bases (3.5 base pai maximal quenching of 88%, and an intrinsic association const of 2×10^7 M^{-1}. For the recombinant Sac7d protein this correspo to a site size of 6.8 bases (3.4 base pairs), maximal quenching 87%, and an association constant of 2×10^7 M^{-1}.

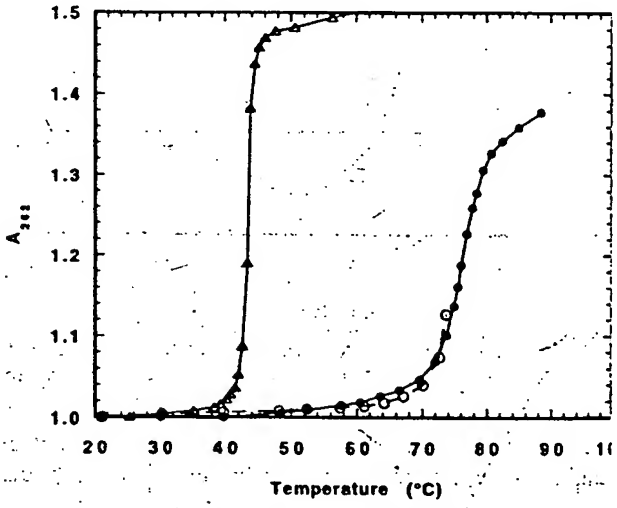


FIGURE 9: Thermal denaturation of poly[dAdT]·poly[dAdT] mo tored by changes in UV absorbance at 262 nm in 0.01 M KH_2PO pH 7.0. The melting of poly[dAdT]·poly[dAdT] is shown alo (open triangles), with native Sac7 proteins (solid circles), and w recombinant Sac7d (open circles). The concentration of pol [dAdT]·poly[dAdT was 70 μM (nucleotides), and the concentrati of protein was 350 μM.

Thermal Stability. Sac7 proteins are highly thermostab as expected from their origin. Native Sac7 and recombina Sac7d samples heated to 100 °C showed no precipitation cloudiness, although some increase in scattering was notic able in the UV spectrum. The proteins unfold reversibly indicated by the observation of similar endotherms wi repetitive DSC scans up to 100 °C.

The native Sac7 proteins show a DSC endotherm at p 6.0 (0.01 M KH_2PO_4, 0.1 M KCl, 0.001 M EDTA) with T_m of 99.0–100.2 °C (data not shown). By comparison, t native Sso7 protein has a T_m of 99.4 °C under simil conditions (data not shown). A precise midpoint for t unfolding transition is difficult to define since data abo

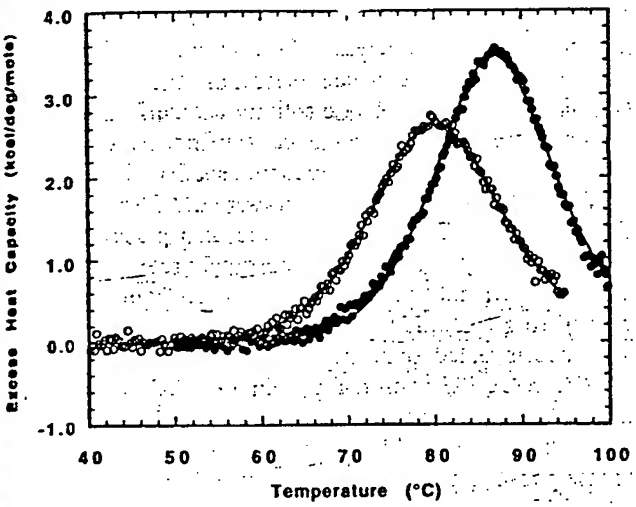


FIGURE 10: Differential scanning calorimetry (DSC) of native Sac7 (solid circles) and recombinant Sac7d (open circles) proteins at pH 4.0 (0.3 M KCl, 0.05 M potassium acetate). Protein concentrations were 1.5 mg/mL of native Sac7 proteins and 1.38 mg/mL of recombinant Sac7d. Smooth curves through the data are nonlinear least-squares fits with T_m = 80.3 °C, ΔH_{cal} = 53.0 kcal/mol, ΔH_{vh} = 49.6 kcal/mol, for the recombinant protein; and T_m = 86.8 °C, ΔH_{cal} = 56.4 kcal/mol, ΔH_{vh} = 60.3 kcal/mol for the native protein.

100 °C cannot be collected in water in the MC2 calorimeter. Notably, the unfolding of the native Sac7 proteins is remarkably reversible, as indicated by essentially 100% reproducibility of successive scans on the same sample following cooling. The recombinant Sac7d protein unfolds at pH 6.0 (0.01 M KH_2PO_4, 0.1 M KCl, 0.001 M EDTA) with a T_m of 92.7 °C, or approximately 7 °C less than the native.

A reliable analysis of the DSC endotherms requires a more complete delineation of the endotherm which can be obtained by lowering the pH and increasing the salt concentration to shift the endotherms to lower temperature. At pH 4.0 (0.05 M potassium acetate, 0.3 M KCl) the native protein unfolds with a T_m of 86.8 °C (Figure 10). The endotherm can be fit with a van't Hoff enthalpy of 60.3 kcal/mol and a calorimetric enthalpy of 56.4 kcal/mol, i.e., a $\Delta H_{cal}/\Delta H_{vh}$ of 0.94, indicating that the native protein exists as a monomer under these conditions and unfolds in an all-or-none fashion with no significant, populated intermediates.

The recombinant Sac7d protein similarly unfolds reversibly at pH 4.0 (0.05 M potassium acetate, 0.3 M KCl) but with a midpoint temperature of 80.3 °C (Figure 10), or 6.5 °C less than the native protein. It unfolds with a van't Hoff enthalpy of 49.6 kcal/mol, and a calorimetric enthalpy of 53.0 kcal/mol, i.e., a $\Delta H_{cal}/\Delta H_{vh}$ of 1.07. The identity, within experimental error, of the calorimetric and van't Hoff enthalpies indicates that the recombinant protein also exists as a monomer under these conditions and unfolds via a two-state reaction.

DISCUSSION

We report here the cloning and sequencing of two genes from *S. acidocaldarius* coding for Sac7 proteins which correspond to Sac7d and Sac7e. The *sac7d* and *sac7e* genes differ at only 16 positions within the coding region (underlined in Figure 3); three of these differences are transversions, while the rest are transitions. The *sac7d* and *sac7e* genes code for 66 and 65 amino acid proteins, respectively. The deduced amino acid sequences are in complete agreement with the published sequences for both proteins (Kimura et al., 1984; Choli et al., 1988a) with the exception of initiator methionines at the amino termini and an additional lysine (Lys66) at the carboxy terminus of the Sac7d protein in the deduced sequence. The additional lysine can be explained either by a failure to discern the final lysine in the amino acid sequencing of the Sac7d or by posttranslational carboxy-terminal processing to produce the mature protein. It should be noted that Sac7d, Sac7e, and Sso7d all terminate with at least two lysine residues (Figure 1).

The data presented here indicate that there are only two Sac7 protein genes in *S. acidocaldarius*. Genes coding for Sac7 proteins other than Sac7d and e could not be detected. The failure to detect genes for the Sac7a and b proteins and the fact that the proteins appear to be simply truncated at the carboxy termini to various extents suggest that Sac7a and b result from either posttranslational modification at the carboxy terminus or by proteolysis during protein isolation and purification.

Promoter elements consistent with the archaeal "A-box" and "B-box" consensus sequences have been located upstream of the *sac7d* and *sac7e* protein coding sequences. The agreement of the "A-box" sequence of *sac7d* with the consensus "A-box" sequence is greater than that for the *sac7e*. This difference between the "A-box" of the promoter elements in the two genes may explain the higher levels of Sac7d relative to Sac7e *in vivo* (Grote et al., 1986).

There is significant sequence similarity in the regions of *sac7d* and *sac7e* extending from the 5′ end of the "A box" to the initiation codon when the corresponding "A-" and "B-" boxes are aligned. The two sequences also have similarly placed pyrimidine rich regions downstream of their termination codons. These regions show similarity to the transcription termination signals described for the *Sulfolobus* virus-like particle, SSV1, where transcription termination has been shown to occur within pyrimidine-rich regions directly 3′ of the consensus TTTTTYT [reviewed in Brown et al. (1989)]. Northern analysis of *S. acidocaldarius* RGJM RNA probed with an oligonucleotide (oligonucleotide F, Table 1) complementary to the common sequence at residues 305–324 of the two *sac7* genes (Figure 3) showed hybridization to a single size of transcripts (Shao and Gupta, unpublished results), indicating that both transcripts terminate in similarly placed regions. Thus, it is likely that the conserved oligopyrimidine sequences of the two genes contain the transcription termination signals.

Although the regions associated with transcription termination are highly homologous, the sequences between these regions and the termination codons are significantly different in the *sac7d* and *sac7e* genes. Similarly, though the regions encompassing the putative core promoter elements in the two genes ("A-" and "B-" boxes) share extensive homology, the sequences 5′ of the "A-box" show less similarity. It would appear that sufficient time has elapsed since the supposed original gene duplication for the two sequences to diverge. The conservation of cis-regulatory elements along with coding regions in the two genes indicates that there is a selective pressure to maintain not only the expression of both gene products but also a large part of their sequence. It is not clear if there is more than one form of the Sso7 proteins.

A typical ribosome binding site sequence upstream of initiator ATG is not observed in either of the two *sac7* genes

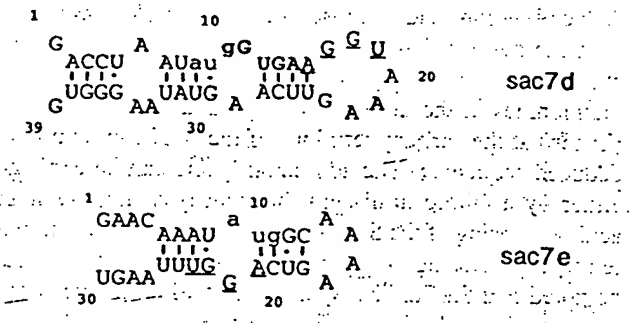


FIGURE 11: Potential secondary structures for the 5′-terminal regions of the *sac7* RNA transcripts determined using Mulfold (Jaeger et al., 1989a,b; Zuker, 1989). Initiator codons are shown in lower case. Putative ribosome binding sequences GGUGA and AGGU are indicated in bold and underlined formats, respectively. Note that the AGGU sequences within the two transcripts are located at different positions.

(Figure 3). This is not unusual, since many other *Sulfolobus* genes also lack these sites (Amils et al., 1993; Dalgaard & Garrett, 1993). However, potential ribosome binding sites are observed downstream of the initiator codons of the two *sac7* genes which have precedents in other archaea. The ribosome binding sites in certain halobacterial genes, which have very short or no 5′ untranslated regions, occur within loops of potential hairpin structures in the 5′ regions of the transcripts (Brown et al., 1989; Amils et al., 1993). The hairpin arrangement probably exposes these sites for interaction with 16S rRNA. We note that the 5′ regions of the two *sac7* transcripts can be folded into secondary structures as shown in Figure 11. The sequence UCACCU near the 3′ end of 16S rRNA of *Sulfolobus* (Woese et al., 1984; Olsen et al., 1985) potentially can either form five base pairs with GGUGA within codons 1−3 or form four base pairs with AGGU within codons 3−4 of the *sac7d* transcript. Corresponding sequences in the sac7e transcript are GGCAA and AAGU, respectively, which cannot form similar pairs with the 16S rRNA. However, further downstream in the *sac7e* transcript, there is AGGU within codons 5−6, which can form four base pairs with the same UCACCU sequence of the 16S rRNA; the corresponding site in sac7d is less efficient AAGU. Parts of these potential ribosome binding sites do occur within single-stranded regions (Figure 11), as are the cases for the above mentioned halobacterial genes. The differences between the sequences and locations of the potential ribosome binding sites of the two *sac7* transcripts, along with the previously mentioned differences in the "A-box" sequences, may also explain the higher synthesis of Sac7d protein.

Kimura et al. (1984) have previously noted that the clustering of lysines in the amino terminus of these proteins is reminiscent of that observed in eukaryotic HMG proteins. Choli et al. (1988b) have also pointed out a slight sequence similarity with E2A DNA-binding protein from adenovirus. An extensive search of the currently available sequence databases showed no significant homologies between the Sac7d protein and any known chromatin or DNA-binding protein. A BLAST search using the Sac7d sequence picked up a 100% homology with the amino-terminal sequence (only 12 amino-terminal residues are known) of a small protein (accession number S21168) from *S. solfataricus* which apparently catalyzes disulfide bond formation (Guagliardi et al., 1992). This report should be viewed with caution due to the loss of activity upon cation exchange chromatography

of th otein. BLAST also picked up a high homolo; a reported p2 ribonuclease (Fusi et al., 1993) fro *solfataricus* with a sequence identical to the Sso7d pr (Choli et al., 1988a). RNase activity for the 7 kDa pro is surprising and remains to be confirmed. Prelimi experiments indicate that the recombinant Sac7d protein not have RNase activity (Edmondson and Shriver, un lished results). The BLAST search also picked up s weak homology with the 30S ribosomal protein S5 fro *coli* (P02356) and heat shock protein X16 from the Afi clawed frog (A22175). A FASTA search using the S sequence revealed some homology with elongation f 1-δ (P29692), 30S ribosomal protein S8 (P24353), and D directed RNA polymerase subunit A′ (P31813). A PROS search using the Sac7d sequence revealed phosphocre kinase phosphorylation sites at residues 17−19 (TSK), 42 (TGR), and 46−48 (SEK), and creatine kinas phosphorylation sites at 33−36 (TYDD), and 46−49 (SE A BLOCKS analysis provided a single meaningful m with ribosomal S5 protein.

We have expressed the *sac7d* gene in the tightly contr BL21(DE3)pLysS *E. coli* expression system develope Studier et al. (1990) using the pET series of plasn Accumulation of the *sac7d* gene product appears to be l in *E. coli*. This is indicated perhaps most clearly by inability to establish the pET-3b/*sac7d* construct ~~BL21~~(DE3). The additional regulation provided by th lysozyme inhibition of T7 polymerase appears to be requ The purified, recombinant protein can be isolated reasonable yield, e.g., typically, about 1 mg of protein p of wet weight *E. coli* cells is obtained, or approximately t that obtained for the native protein from *S. acidocalda* We have been unsuccessful in expressing the *sac7e* g possibly due to its usage of codons rare in *E. coli*.

The recombinant Sac7d protein appears to be essenti identical to the native Sac7 proteins in all respects ex for stability. The UV spectral extinction coefficients identical, as are the fluorescence excitation and emis spectra. This is perhaps not surprising given that both largely due to a single tryptophan on the surface of protein (Edmondson, Qiu, and Shriver, manuscript submit [see also Baumann et al. (1994) for the structure of Sso although the two tyrosines should be sensitive to differer in structure. CD spectra are more sensitive to differer in secondary structure content, and the spectra of the proteins are essentially identical, again indicating sim structures for native and recombinant protein.

Analyses of the CD spectra using the variable selec method of Johnson (Manavalan & Johnson, 1987) indi that Sac7d consists of 31% helix and 22−25% β-sheet. differs from the 52% α-helix, 12% β-sheet predicted sequence analysis algorithms in this work and the 1 α-helix, 15% β-sheet predicted by Choli et al. (1988a) u the average of four different prediction methods. All of th methods significantly underestimate the amount of β-sl in Sac7d (42%) as determined from the NMR solu structure (Edmondson, Qiu, and Shriver, manuscript sub ted) [see also Baumann et al. (1994)]. However, the hel content determined by CD (31%) is close to that of the N solution structure (22% α-helix, 11% 3_{10}-helix). An anal of the CD spectrum of Sac7e (Dijk & Reinhardt, 1986) us the PG method (Provencher & Glockner, 1981) gave a m better estimate of β-sheet content (44%) but underestima

the helical content (15%). The CD spectrum reported for Sac7e (Dijk & Reinhardt, 1986) differs quantitatively from that of native Sac7 and recombinant Sac7d presented here. Further, the inability of the CD analyses to accurately estimate the secondary structure content suggests that at least part of the secondary structure contributions to the CD spectra of the Sac7 proteins are not well represented in these sets of reference proteins.

A more detailed, atomic level comparison of the structures of the recombinant and native proteins can be obtained from NMR. The "fingerprint" region of double-quantum filtered COSY spectra of proteins shows the chemical shift correlations of alpha and NH protons and is exquisitely sensitive to the structure of the protein [see, for example, Wishart et al. (1992)]. This permits a qualitative comparison of the structure of the backbone of the two proteins which is more detailed than that provided by optical spectra comparisons. The fingerprint regions of native and recombinant Sac7d protein are remarkably similar, indicating that the two proteins have very similar backbone folding patterns.

The binding of the Sac7 proteins to double stranded DNA leads to a dramatic decrease in intrinsic tryptophan fluorescence. The large signal allows for essentially noise-free titrations and accurate comparisons of the native and recombinant protein binding function. The data presented here indicate an affinity of 2×10^7 M^{-1} and site size of 3.5 base pairs for poly[dGdC]·poly[dGdC]. The agreement of quantitative binding parameters obtained for the native and recombinant proteins is additional evidence for essentially identical global folds for the two proteins. These binding studies are the first quantitative analysis of the binding of the Sac7 proteins to DNA.

Various prior studies of the 7 kDa DNA-binding proteins from *Sulfolobus* have characterized the binding to nucleic acids in a qualitative manner. Electron micrographs of the 7 kDa proteins from *S. acidocaldarius* complexed with DNA indicated that the helix becomes increasingly compacted with increasing ratios of protein to DNA (Dijk & Reinhardt, 1986; Lurz et al., 1986). Filter binding studies confirmed that the 7 kDa proteins had an affinity for pBR322 DNA even at relatively high salt concentrations (e.g., 0.265 M NaCl) which was comparable to that observed for *E. coli* HU protein (Grote et al., 1986; Choli et al., 1988a). Characterization of the affinity for DNA in this work was in terms of percent bound at a specific ratio of protein to DNA. DNA-melting studies have also been performed on a small DNA-binding protein from *S. acidocaldarius*, HSNP-C', with an amino acid composition similar to the Sac7e protein, although the sequence has not been presented The protein increases the T_m of double-stranded DNA (Reddy & Suryanarayana, 1989). In addition, this protein demonstrated a significant quenching of its intrinsic tryptophan fluorescence upon DNA binding, although no quantitative analysis of the titrations was performed.

Baumann et al. (1994) have recently presented some fluorescence binding data for the homologous Sso7 proteins from *S. solfataricus*. A quantitative analysis of the titrations was not performed, but a visual inspection of the data indicates a binding site size for double-stranded DNA of six base pairs in low salt (0.02 M Tris, pH 7.4), nearly twice that presented here. Assuming a site size of 3−6 base pairs, the binding affinity in low salt is approximately 0.5 to 1 × 10^6 M^{-1}. The thermal stability of poly[dIdC]·poly[dIdC] was increased by approximately 40 °C in 5 mM Tris (pH 7.0).

The unfolding of both the native and recombinant proteins is reversible, allowing for detailed, accurate characterization of the thermodynamics of folding. In contrast to all other physical parameters studied here, the energetics of folding of the recombinant Sac7d protein differs significantly from that of the native Sac7 proteins. The native protein unfolds at pH 6.0 at 100 °C, remarkable given the absence of any metal cofactors or disulfides. Surprisingly, the recombinant protein unfolds with a T_m 6.5 °C less than the native. The lower enthalpy of unfolding of the recombinant protein is not surprising and most likely results from a positive heat capacity change associated with unfolding. Any shift to lower temperature of an endotherm associated with a positive ΔC_p will lead to a decrease in enthalpy since

$$\Delta C_p = \left(\frac{\partial \Delta H}{\partial T}\right)_P$$

It is generally thought that a positive ΔC_p of unfolding is due to the exposure of internal hydrophobic residues (Sturtevant, 1977; Privalov & Gill, 1988). The magnitude of the change observed here is consistent with that observed for other globular proteins (Privalov & Gill, 1988).

Maras et al. (1992) have previously noted that specific lysine monomethylation of glutamate dehydrogenase from *S. solfataricus* might be responsible for enhanced thermal stability of this enzyme relative to homologous mesophile forms. Baumann et al. (1994) have presented mass spectroscopic evidence correlating methylation of the Sso7 protein with growth temperature, and they have suggested that such a modification might be related to the stability of the protein. The most straightforward way to determine if methylation increases the thermostability of the protein would be to compare the stabilities of the protein in its methylated and unmethylated forms. Demethylation of the native protein is not a trivial control experiment given the lack of commercially available demethylases and most importantly the specificity of reported demethylases (Paik & Kim, 1980). In the absence of a demethylase, the preparation of an unmethylated form is best accomplished using recombinant protein. We have demonstrated here a significant difference in the thermostability of native and recombinant Sac7 protein. The only known difference between these proteins is the ϵ-aminomonomethylation of lysines 5 and 7 in the native protein and the initiating methionine in the recombinant protein. The lack of Lys66 in the reported amino acid sequence of the native protein is presumably a sequencing error, and this will be investigated in the NMR analysis of the native protein. No other posttranslational modification, such as phosphorylation or glycosylation, of the native or recombinant Sac7 proteins was detectable. The current evidence, therefore, strongly indicates that *Sulfolobus* can increase the thermostability of some of its proteins by specific lysine monomethylation.

We note that the level of specific methylation of Sac7 is variable and incomplete, i.e., the native preparation is heterogeneous (Kimura et al., 1984; Choli et al., 1988a,b). Choli et al., (1988b) report that the degree of monomethylation of lysine 4 is 70%, 25%, and 20% in native Sac7a, Sac7b, and Sac7d, respectively; while that for lysine 6 is 50%, 40%, and 50%, respectively. Heterogeneity would be

expected to lead to broadening of the endotherm, rather than narrowing (see Figure 10). It would appear, therefore, that stabilization might not require complete methylation of the specific lysines.

Interestingly, we have been unable to increase the stability of the recombinant Sac7d protein by nonspecific, reductive methylation (McCrary and Shriver, unpublished results), a process which leads to predominantly dimethylation (Means & Feeney, 1971). Monomethylation changes the pK_a of the ϵ-amino group from 9.25 to 10.63, while dimethylation has little further effect giving a pK_a of 10.78 (Paik & Kim, 1980). Trimethylation returns the pK_a to 9.8. Given the small change in pK_a and the fact the difference is observed even at pH 4.0, it is doubtful that an effect of monomethylation on stability might be electrostatic in origin. A structural explanation of the difference in stability must await a more detailed comparison of the structures of the native and recombinant proteins. The spectroscopic data presented here would indicate that the structural differences are slight.

ACKNOWLEDGMENT

We thank Drs. Paul Hargrave and Hugh McDowel (University of Florida) for amino-terminal sequencing of the 7 kDa proteins of *Sulfolobus acidocaldarius*, F. William Studier (Brookhaven) for providing the expression vectors and *E. coli* strains HMS174 and BL21 and their derivatives, Jack Parker (Southern Illinois University) for supplying *E. coli* CJ236, Dennis Grogan (University of Cinncinati) for providing *S. acidocaldarius* DG6 and *S. solfataricus* P2 and advice for characterization of different *Sulfolobus* strains, and Lingshi Qiu (Southern Illinois University) for purifying the Sac7 and Sso7 proteins from *Sulfolobus* strains. We also thank Dr. Ignatius Gomes for valuable discussions and Neelima Reddy for her assistance in the laboratory. Finally, we thank an anonymous reviewer for pointing out the GGTGA sequence as a potential ribosome binding site.

REFERENCES

Altshul, S., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.

Amils, R., Cammarano, P., & Londei, P. (1993) in *The Biochemistry of Archaea (Archaebacteria)* (Kates, M., Kushner, D. J., & Matheson, A. T., Eds.) pp 393−438, Elsevier, New York.

Bairoch, A. (1992) *Nucleic Acids Res. 20*, 2013−2018.

Baumann, H., Knapp, S., Lundback, T., Ladenstein, R., & Hard, T. (1994) *Struct. Biol. 1*, 808−819.

Berger, L. S., & Kimmel, A. R. (1987) *Guide to Molecular Cloning Techniques*, Academic Press, New York.

Brock, T., Brock, K., Belly, R., & Weiss, R. (1972) *Arch. Microbiol. 84*, 54−68.

Brown, J., Daniels, C., & Reeve, J. (1989) *CRC Crit. Rev. Microbiol. 16*, 287−338.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285−298.

Chen, G. C., & Yang, J. T. (1977) *Anal. Lett. 10*, 1195.

Choli, T., Henning, P., Wittmann-Liebold, B., & Reinhardt, R. (1988a) *Biochim. Biophys. Acta 950*, 193−203.

Choli, T., Wittmann-Liebold, B., & Reinhardt, R. (1988b) *J. Biol. Chem. 263*, 7087−7093.

Chou, P. Y., & Fasman, G. (1974) *Biochemistry 13*, 222−245.

Chou, P. Y., & Fasman, G. (1978) *Annu. Rev. Biochem. 47*, 251−276.

Dalgaard, J. Z., & Garrett, R. A. (1993) in *The Biochemistry of Archaea (Archaebacteria)* (Kates, M., Kushner, D. J., & Matheson, A. T., Eds.) pp 535−563, Elsevier, New York.

Debois, D., Gilles, K., Hamilton, J., Rebers, P., & Smith., F. (1 *Anal. Chem. 28*, 350.

DeRosa, M., & Gambacorta, A. (1975) *J. Gen. Microbiol. 86*, 1 164.

Dijk, J., & Reinhardt, R. (1986) in *Bacterial Chromatin* (Gua C. O., & Pon, C. O., Eds.) Springer-Verlag, Berlin.

Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.

Emory, S. A., & Belasco, J. G. (1990) *J. Bacteriol. 172*, 44 4481.

Fiske, C., & Subbarow, Y. (1925) *J. Biol. Chem. 66*, 375−4

Fuchs, R. (1991) *Comput. Appl. Biosci. 7*, 105−106.

Fusi, P., Tedeschi, G., Aliverti, A., Ronchi, S., Tortora, P Guerritore, A. (1993) *Eur. J. Biochem. 211*, 305−310.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. 120*, 97−120.

Gill, S., & von Hippel, P. (1989) *Anal. Biochem. 182*, 319−

Gordon, A., & Ford, R. (1972) *The Chemist's Companion Handbook of Practical Data, Techniques, and References*, Wiley, New York.

Grodberg, J., & Dunn, J. J. (1988) *J. Bacteriol. 170*, 1245−1

Grogan, D. (1989) *J. Bacteriol. 171*, 6710−6719.

Grogan, D. W. (1991) *J. Bacteriol. 173*, 7725−7727.

Grote, M., Dijk, J., & Reinhardt, R. (1986) *Biochim. Biophys. 873*, 405−413.

Guagliardi, A., Cerchia, L., De Rosa, M., Rossi, M., & Bartol S. (1992) *FEBS Lett. 303*, 27−30.

Gupta, R. (1984) *J. Biol. Chem. 259*, 9461−9471.

Henikoff, S., & Henikoff, J. G. (1991) *Nucleic Acids Res. 19*, 65 6572.

Hirs, C. W. (1967) *Methods Enzymol. 11*, 411−413.

Inman (1962) *J. Mol. Biol. 5*, 172.

Jaeger, J. A., Turner, D. H., & Zuker, M. (1989a) *Proc. Natl. A Sci. U.S.A. 86*, 7706−7710.

Jaeger, J. A., Turner, D. H., & Zuker, M. (1989b) *Methods Enzy 183*, 281−306.

Johnson, R., & Walseth, T. F. (1979) *Adv. Cyclic Nucleotide 10*, 135−167.

Johnson, W. C., Jr. (1984) *Food Analysis Principles and Te niques, Volume 2, Physicochemical Techniques*, (Gruenwe D. W., & Whitaker, J. R., Eds.) Marcel Dekker, Inc., New Y

Kimura, M., Kumura, J., Phillips, D., Reinhardt, R., & Dijk (1984) *FEBS Lett. 176*, 176−178.

Kunkel, T., Roberta, J. D., & Zakour, R. A. (1987) *Meth Enzymol. 154*, 367−382.

Leloir, L., & Cardini, C. (1957) *Methods Enzymol. 3*, 840−8

Lurz, R., Grote, M., Dijk, J., Reinhardt, R., & Dobrinski, B. (19 *EMBO J. 5*, 3715−3721.

Manavalan, P., & Johnson, W. C. (1987) *Anal. Biochem. 167*, 7 85.

Maras, B., Conslavi, V., Chiaraluce, R., Politi, L., De Rosa, Bossa, F., Scandurra, R., & Barra, D. (1992) *Eur. J. Biochem 203*, 81−87.

Mayes, E. L. V. (1984) in *Methods in Molecular Biology, Volu 1: Proteins* (Walker, J. M., Ed.) Humana Press, Clifton, NJ

McAfee, J. (1993) Ph.D. Dissertation, Genetic and DNA Bindi Studies on a 7 kDa Protein from *Sulfolobus acidocaldari* Southern Illinois University, Carbondale, IL.

McGhee, J., & von Hippel, P. (1974) *J. Mol. Biol. 86*, 469−4

Means, G., & Feeney, R. (1971) *Chemical Modification of Protei* Holden-Day, Inc., San Francisco, CA.

Moore, S., & Stein, W. H. (1954) *J. Biol. Chem. 211*, 907−9

Olsen, G. J., Pace, N. R., Nuell, M., Kaine, B. P., Gupta, R., Woese, C. R. (1985) *J. Mol. Evol. 22*, 301−307.

Paik, W., & Kim, S. (1980) *Protein Methylation*, John Wiley a Sons, New York.

Privalov, P., & Gill, S. (1988) *Adv. Protein Chem. 39*, 191−2

Provencher, S. W., & Glöckner, J. (1981) *Biochemistry 20*, 3 37.

Rance, M., Bodenhausen, G., Wagner, G., Ernst, R., & Wüthric K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.

Reddy, T. R., & Suryanarayana, T. (1988) *Biochim. Biophys. Ac 949*, 87−96.

Reddy, T. R., & Suryanarayana, T. (1989) *J. Biol. Chem. 26* 17298−17308.

Robson, B., & Suzuki, E. (1976) *J. Mol. Biol. 107*, 327−356.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sandman, K., Krzycki, J. A., Dobrinski, B., Lurz, R., & Reeve, J. N. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5788–5791.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Savitsky, A., & Golay, M. J. E. (1964) *Anal. Chem. 36*, 1627.
Schägger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368–379.
Searcy, D. G. (1975) *Biochim. Biophys. Acta 395*, 535–547.
Searcy, D. G., & Delange, R. J. (1980) *Biochim. Biophys. Acta 609*, 197–200.
Short, J., Fernandez, J. M., Sorge, J. A., & Huse, W. D. (1988) *Nucleic Acids Res. 16*, 7583–7600.
Shriver, J. W., & Kamath, U. (1990) *Biochemistry 29*, 2556–2564.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Stein, D. B., & Searcy, D. G. (1978) *Science 202*, 219–221.
Studier, W. F., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 66–89.
Sturtevant, J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236–2240.
Thomm, M., Stetter, K. O., & Zillig, W. (1982) *Zbl. Bakt. Hyg., I. ABT. Orig. C3*, 128–139.
Tiktopulo, E. I., & Privalov, P. (1974) *Biophys. Chem. 1*, 349–357.
van Iersel, J., Jzn, J. F., & Duine, J. (1985) *Anal. Biochem. 151*, 196–204.
Wells (1970) *J. Mol. Biol. 54*, 465.
Wishart, D., Sykes, B., & Richards, F. (1992) *Biochemistry 31*, 1647–1651.
Woese, C. R., Gupta, R., Hahn, C. M., Zillig, W., & Tu, J. (1984) *System. Appl. Microbiol. 5*, 97–105.
Zillig, W. (1993) *Nucleic Acids Res. 21*, 5273.
Zillig, W., Palm, P., Reiter, W.-D., Gropp, F., Puhler, G., & Klenk, H.-P. (1988) *Eur. J. Biochem. 173*, 473.
Zuker, M. (1989) *Science 244*, 48–52.
Zuiderweg, E. R. P., Hallenga, K., & Olejniczak, E. T. (1986) *J. Magn. Reson. 70*, 336–343.

BI950704Z